Science

4 July 2008 | $10

MESSENGER
Returns to Mercury

AAAS

CONTENTS

Volume 321, Issue 5885

COVER

A mosaic of visible to near-infrared images of the surface of Mercury, obtained by the MESSENGER spacecraft on 14 January 2008. The circular feature in the upper right is the Caloris impact basin, 1500 kilometers in diameter. Results from the flyby are discussed in a special section beginning on page 58.

Image: NASA/Johns Hopkins University Applied Physics Laboratory/Arizona State University/ Carnegie Institution of Washington

DEPARTMENTS

11 *Science* Online

12 This Week in *Science*

16 Editors' Choice

18 Contact *Science*

19 Random Samples

21 Newsmakers

144 New Products

145 *Science* Careers

EDITORIAL

15 Reviewing Peer Review
by Bruce Alberts, Brooks Hanson, Katrina L. Kelner
>> Letter by M. Raff et al. *p. 36*

SPECIAL SECTION

MESSENGER

INTRODUCTION

Mercury, Up-Close Again 58

REPORTS

Return to Mercury: A Global Perspective on MESSENGER's First Mercury Flyby 59
S. C. Solomon et al. >> Science *Podcast*

Spectroscopic Observations of Mercury's Surface Reflectance During MESSENGER's First Mercury Flyby 62
W. E. McClintock et al.

Reflectance and Color Variations on Mercury: Regolith Processes and Compositional Heterogeneity 66
M. S. Robinson et al.

Volcanism on Mercury: Evidence from the First MESSENGER Flyby 69
J. W. Head et al.

Geology of the Caloris Basin, Mercury: A View from MESSENGER 73
S. L. Murchie et al.

Laser Altimeter Observations from MESSENGER's First Mercury Flyby 77
M. T. Zuber et al.

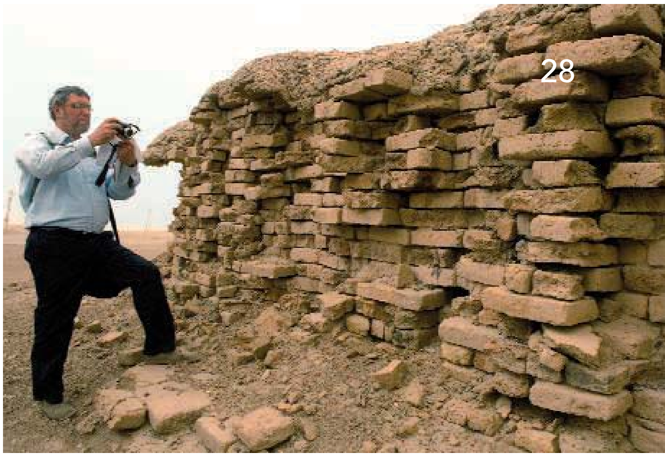
Mercury Cratering Record Viewed from MESSENGER's First Flyby 79
R. G. Strom et al.

The Structure of Mercury's Magnetic Field from MESSENGER's First Flyby 82
B. J. Anderson et al.

Mercury's Magnetosphere After MESSENGER's First Flyby 85
J. A. Slavin et al.

MESSENGER Observations of the Composition of Mercury's Ionized Exosphere and Plasma Environment 90
T. H. Zurbuchen et al.

Mercury's Exosphere: Observations During MESSENGER's First Mercury Flyby 92
W. E. McClintock et al.



NEWS OF THE WEEK

Big Land Purchase Triggers Review of Plans to Restore Everglades 22

California Emissions Plan to Explore Use of Offsets 23

Life's Innovations Let It Diversify, at Least Up to a Point 24
>> Research Article p. 97

Where Have All Thoreau's Flowers Gone? 24

SCIENCESCOPE 25

Billion-Dollar Cancer Mapping Project Steps Forward 26

Panel to Take Closer Look at Scientific Whaling 26

Tiny Transistor Gets a Good Sorting Out 27
>> Report p. 101

NEWS FOCUS

Preserving Iraq's Battered Heritage 28
On Iraq's Front Lines
>> Science Podcast

Epilepsy: When Death Strikes Without Warning 31

Competing Teams Plot Two Different Paths to a New Particle Smasher 34

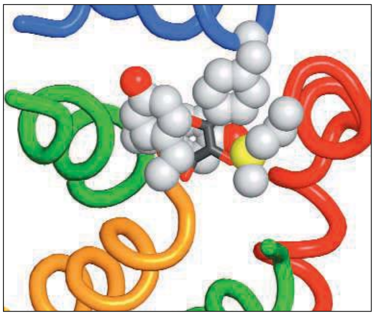
CONTENTS continued >>

SCIENCE EXPRESS

www.sciencexpress.org

MOLECULAR BIOLOGY
A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome
M. Sultan et al.
Shotgun sequencing of 27–base pair segments of messenger RNA from human kidney and immune cells identifies previously undescribed transcriptional units and splice junctions.
10.1126/science.1160342

ATMOSPHERIC SCIENCE
BREVIA: N_2O_5 Oxidizes Chloride to Cl_2 in Acidic Atmospheric Aerosol
J. M. Roberts, H. D. Osthoff, S. S. Brown, A. R. Ravishankara
Laboratory studies affirm that the oxidation of chloride ions in aerosols by N_2O_5 is a significant source of chlorine in the troposphere, a major reactant that helps form ozone.
10.1126/science.1158777



BIOCHEMISTRY
The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na^+/Sugar Symport
S. Faham et al.
The structure of a sugar transporter suggests how these proteins may rearrange to permit the sugar to enter and leave the binding site on opposite sides of the membrane.
10.1126/science.1160406

LETTERS

Painful Publishing *M. Raff, A. Johnson, P. Walter* 36
>> *Editorial p. 15*
The Enemy Within *V. L. Roggli*
The Limits of Water Pumps *L. Wang and P. D'Odorico*
Omissions in GLAST Story *G. F. Bignami* et al.
Life in Science: Frogs on a Plane *J. Rigg* 37

BOOKS *ET AL.*

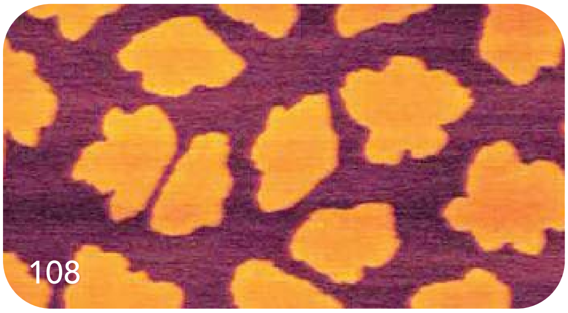
Gorilla Mountain *R. Ebersole;* **Robo World** *J. D. Brown;* **Nature's Machines** *F. Watts; reviewed by S. Kovats* 39
Ghostwalk *R. Stott, reviewed by J. Golinski* 40
Falling for Science Objects in Mind *S. Turkle, Ed.* 41
The City and the Stars *A. C. Clarke* 42
BROWSINGS 42

POLICY FORUM

An Earth Systems Science Agency 44
M. Schaefer et al.

PERSPECTIVES

The Scale of Experience 46
M. E. Hasselmo >> *Report p. 140*
Indirect Social Influence 47
J. Denrell
Transient Dynamics for Neural Processing 48
M. Rabinovich, R. Huerta, G. Laurent



108

PERSPECTIVES *CONTINUED...*

A Unique Platform for Materials Design 50
T. P. Lodge
Carbon Emissions and Acidification 51
R. E. Zeebe, J. C. Zachos, K. Caldeira, T. Tyrrell

REVIEW

NEUROSCIENCE
Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations 53
T. Klausberger and P. Somogyi

41

BREVIA

EVOLUTION
Reduced Responses to Selection After Species Range Expansion 96
B. Pujol and J. R. Pannell
An annual spurge with a geographical range that expanded after the Ice Age shows decreased response to selection at the edges of its new range, as predicted by theory.

RESEARCH ARTICLE

PALEONTOLOGY
Phanerozoic Trends in the Global Diversity of Marine Invertebrates 97
J. Alroy et al.
A compilation of more than 3 million specimens of fossil marine invertebrates shows that their diversity increased more in the Jurassic and less since then than an earlier study implied.
>> *News story p. 24*

REPORTS

APPLIED PHYSICS
Self-Sorted, Aligned Nanotube Networks for Thin-Film Transistors 101
M. C. LeMieux et al.
Treating silicon substrates with a silane layer allows them to absorb spin-coated, semiconducting carbon nanotubes, yielding aligned and densely packed nanotube films.

CONTENTS continued >>

REPORTS *CONTINUED...*

ASTRONOMY
Relativistic Spin Precession in the Double Pulsar 104
R. P. Breton et al.
Four years of data track the spin precession of a pulsar orbiting a second pulsar, providing a positive test of general relativity in a strong gravitational field.

MATERIALS SCIENCE
Characterization of Step-Edge Barriers in Organic Thin-Film Growth 108
G. Hlawacek et al.
In contrast to the growth of inorganic films, bending of a rod-shaped organic molecule at step edges and its anisotropy leads to a change from growth of layers to terraced mounds.

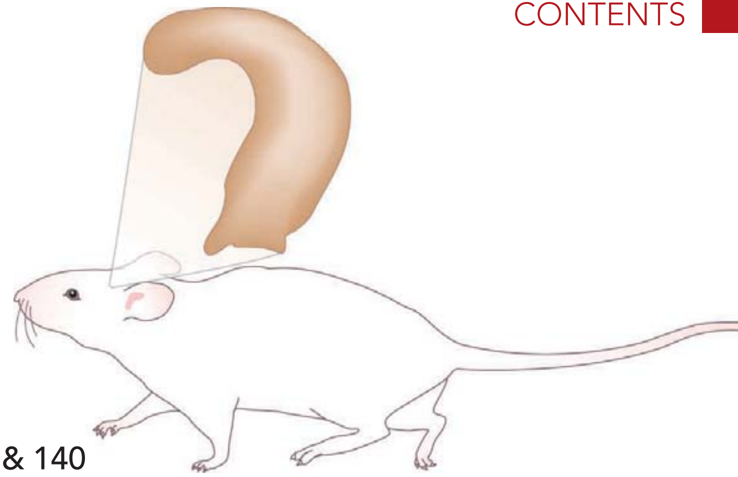
CLIMATE CHANGE
Large and Rapid Melt-Induced Velocity Changes in the Ablation Zone of the Greenland Ice Sheet 111
R. S. W. van de Wal et al.
Measurements of ice velocity across western Greenland show that the ice sheet responds within days to excess meltwater, although annual flow has slowed a bit over 17 years.

CHEMISTRY
Mg/Al Ordering in Layered Double Hydroxides Revealed by Multinuclear NMR Spectroscopy 113
P. J. Sideris, U. G. Nielsen, Z. Gan, C. P. Grey
Rapid sample spinning during nuclear magnetic resonance spectroscopy reveals a highly ordered cation distribution in layered materials.

DEVELOPMENTAL BIOLOGY
Autophagy Is Essential for Preimplantation Development of Mouse Embryos 117
S. Tsukamoto et al.
As fertilized mouse eggs develop into embryos and maternal proteins are eliminated, the degradative process of autophagy is required for proper growth.

EVOLUTION
Phylogenetic Signal in the Eukaryotic Tree of Life 121
M. J. Sanderson
A survey of sequences in GenBank, which represent about 10 percent of described species, shows that the patchy distribution of data is insufficient to build a eukaryotic tree of life.

ECOLOGY
Accelerated Human Population Growth at Protected Area Edges 123
G. Wittemyer et al.
Contrary to expectations, human populations living near protected areas in 45 countries in Africa and Latin America are increasing nearly twice as fast as other rural ones.



46 & 140

CELL BIOLOGY
Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops 126
T. Y.-C. Tsai et al.
Analysis of known and theoretical oscillatory circuits in cells shows that those with both negative and positive feedback are more robust and allow frequency control independent of amplitude.

MEDICINE
Sporadic Autonomic Dysregulation and Death Associated with Excessive Serotonin Autoinhibition 130
E. Audero et al.
In young mice, expression of higher than normal levels of a type of serotonin receptor causes sporadic death with features reminiscent of sudden infant death syndrome. >> Science *Podcast*

BIOCHEMISTRY
Myosin I Can Act As a Molecular Force Sensor 133
J. M. Laakso, J. H. Lewis, H. Shuman, E. M. Ostap
Myosin I, a motor protein that plays a role in hearing, is a sensitive tension sensor, reacting to small loads (less than 2 picoNewtons) by binding for much longer times to actin.

NEUROSCIENCE
The Spread of Ras Activity Triggered by Activation of a Single Dendritic Spine 136
C. D. Harvey, R. Yasuda, H. Zhong, K. Svoboda
When strengthened, individual synapses on dendritic spines contain an activated small regulatory protein that spreads to nearby spines, possibly altering their sensitivity.

NEUROSCIENCE
Finite Scale of Spatial Representation in the Hippocampus 140
K. B. Kjelstrup et al.
The rat hippocampus provides a representation of the animal's entire spatial environment, coding distances up to 1 meter away in the dorsal region and up to 15 meters at the ventral tip.
>> Perspective p. 46

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals Mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2008 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $144 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $770; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery) $85. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**

Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $10.00 current issue, $15.00 back issue prepaid includes surface postage; bulk rates on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $20.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

CONTENTS continued **>>**

Letting go.

SCIENCE NOW

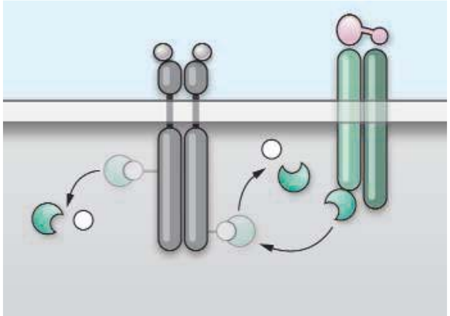
www.sciencenow.org

HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

Why It's Hard to Say Goodbye
Study links loss of a loved one to the brain's pleasure center.

Don't Judge a Plant by Its Species
An ant, an aphid, and a milkweed are changing thoughts about community ecology.

African Lion-Killer Had Help
Virus conspired with tick-borne parasites and extreme droughts.



TIMP-dependent dephosphorylation.

SCIENCE SIGNALING

www.sciencesignaling.org

THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

REVIEW: Tissue Inhibitors of Metalloproteinases in Cell Signaling—Metalloproteinase-Independent Biological Activities
W. G. Stetler-Stevenson
TIMPs can act directly through cell surface receptors or indirectly through modulation of proteases.

PERSPECTIVE: The Cytoplasmic Tail of MUC1—A Very Busy Place
D. D. Carson
The cytoplasmic domain of mucin 1 (MUC1) plays numerous roles in intracellular signaling pathways.

Farewell, Micella.

SCIENCE CAREERS

www.sciencecareers.org/career_development

FREE CAREER RESOURCES FOR SCIENTISTS

Educated Woman, Postdoc Edition, Chapter 18: End of the Road
M. P. DeWhyse
Micella Phoenix DeWhyse celebrates her Independence Day—and we're sad; with related podcast interview.

Taken for Granted: By the Numbers
B. L. Benderly
A committee calls for better government data collection about jobs for scientists.

In Person: Research in France
A. Bikfalvi
Understanding the French public research system is critical to foreign scientists coming into the country.

July 2008 Funding News
J. Fernández
Learn about the latest in research funding, scholarships, fellowships, and internships.

SCIENCE PODCAST

www.sciencemag.org/multimedia/podcast

FREE WEEKLY SHOW

Download the 4 July *Science* Podcast to hear about MESSENGER's first Mercury flyby, possible clues to sudden infant death syndrome, preserving Iraqi antiquities, and more.

CREDITS: (SCIENCE NOW) A. INDEN/ZEFA/CORBIS; (SCIENCE CAREERS) AXEL-D/CREATIVE COMMONS; (SCIENCE SIGNALING) CHRIS BICKEL

Separate individual or institutional subscriptions to these products may be required for full-text access.

EDITED BY STELLA HURTLEY

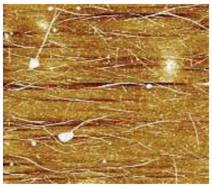
The Human Honeypot >>

The possibility that international investment in protected areas would turn parks into magnets for human immigration (the "honeypot" hypothesis) and thereby reduce conservation effectiveness has been a concern of conservationists, economists, and the development community for some years. **Wittemyer *et al.*** (p. 123) now confirm that rates of human population growth around 306 protected areas in 45 countries across Africa and Latin America are nearly twice the country averages. The high population growth around protected areas is correlated with international donor funding to parks and the consequent creation of park-related jobs and services and, disappointingly, is associated with accelerated rates of deforestation.

Reassessing Past Diversity

Assessing past diversity of life and how it changed over time requires assembly of a database of the many individual and diverse studies of fossils. An early effort was by Jack Sepkoski, and his data on first and last occurrences of marine invertebrates showed an increase in diversity following the Cambrian explosion and particularly since about 100 million years ago. **Alroy *et al.*** (p. 97; see the news story by **Kerr**) have now analyzed a new compilation of more than 3 million specimens resolved to the genus and species level. In contrast to older analyses, the data support a Jurassic increase and imply that the increase in diversity in the Cenozoic was not particularly high relative to earlier times.

Transistor Nanofilms



For single-walled carbon nanotubes to find use in electronics, a method is needed to extract the semiconducting ones from the metallic ones and to deposit only the semiconductors into dense, aligned patterns on a substrate. While these steps have been accomplished individually or on a small scale, the methods are not amenable to large-scale fabrication. **LeMieux *et al.*** (p. 101; see the news story by **Service**) show that by treating silicon substrates with silane monolayers, which selectively absorb semiconducting carbon nanotubes, they can spin-coat solutions of carbon nanotubes to produce films where the nanotubes are aligned and densely packed. The films show excellent transistor behavior including on/off ratios above 100,000.

Test of the Double Pulsar

Under general relativity and strong gravity, when massive objects orbit each other closely, their spin and orbital angular momentum should couple. This coupling leads to precession of the spin of each body. This and other related effects have been difficult to test because it requires close observations of the orientation and spin of two massive objects. **Breton *et al.*** (p. 104) show that the recently discovered double pulsar provides such a test. The geometry of the system is such that one pulsar eclipses the other when viewed from Earth, blocking some of the radio emissions from the companion, which provides information on its orientation and rotation. Four years of data confirm the relativistic coupling and provide a new test of general relativity and strong gravity theories.

Steps in Organic Film Growth

An effect often observed in the growth of inorganic thin films is that atoms are more likely to climb step edges between layers than descend. When this effect is seen, the activated nature of descending the step edge is described with a term called the Ehrlich-Schwoebel barrier. **Hlawacek *et al.*** (p. 108) explored whether such barriers are at work in the more complex growth of molecules used in organic electronics and light-emitting diodes, in this case, the rodlike molecule *para*-sexiphenyl. An analysis of atomic force microscopy images for different film thicknesses revealed the presence of a 0.67 electron volt barrier, and that this barrier relaxes when the layers first start to grow. Thus, the film growth changes from layer-by-layer growth to the formation of terraced mounds. Calculations of the transition state suggest the molecule undergoes bending at the step edge, an effect distinct from those seen in atomic growth.

Water and Ice

Water derived from melting on the surface can be transferred quickly to the base of an ice sheet, lubricating the ice-ground interface and facilitating movement of the ice sheet. How much will global warming accelerate the decay of the polar ice sheets and, consequently, the rate of sea-level rise? **Van de Wal *et al.*** (p. 111) present 16 years of data from the Western margin of the Greenland Ice Sheet and show a correlation between meltwater production and ice velocity on a weekly time scale but observe no sign of a positive feedback over an annual time scale. This suggests that the internal drainage system of the ice sheet adjusts to increases in meltwater inputs, and that annual velocities are mainly functions of ice thickness and surface slope.

Toward the Tree of Life

By quantifying the distribution of phylogenetically informative data across the entire eukaryotes, **Sanderson** (p. 121) has tackled the problem of reconstructing the complete tree of life. The available data are distributed across virtually all groups of eukaryotes, with about 10% of known species represented, but the distribution of the most informative collections of sequences is patchy. Not surprisingly, most of the information-dense clades are among the charismatic megabiota, such as mammals and other vertebrates and flowering plants, with other rich pockets of data centered on experimental model

CREDITS (TOP TO BOTTOM): G. WITTEMYER; LEMIEUX *ET AL.*

Published by AAAS

organisms. A coordinated sampling effort designed to bridge the gaps in phylogenetic information in eukaryotes is now needed, particularly in diverse but poorly studied groups.

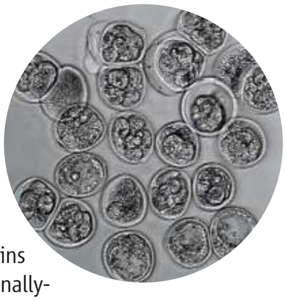
A Mechanism for Sudden Infant Death Syndrome?

Deficits in serotonin neurotransmission have been hypothesized to be involved in sudden infant death syndrome (SIDS), the leading cause of death during the first year of life. **Audero *et al.*** (p. 130) describe a sporadic death phenotype in mice with increased serotonin autoinhibition as a result of overexpression of the serotonin 1A autoreceptor (Htr1a). Deficient serotonergic feedback regulation is sufficient to precipitate autonomic crisis and death. Until now, most SIDS research has focused on respiratory or cardiovascular deficits. These new findings, however, suggest that SIDS is associated with a widespread loss of sympathetic tone, including both bradycardia (slow heart rate) and hypothermia.

Spreading the Word

Individual dendritic spines, the receiving ends of synapses, compartmentalize small diffusible molecules. In particular, Ca^{2+} signals in spines are synapse-specific. However, synapses interact in subtle ways through diffusible postsynaptic factors, which suggests the existence of molecular signals that are activated at individual synapses but that can spread to other synapses. **Harvey *et al.*** (p. 136, published online 12 June) used two-photon glutamate uncaging to induce long-term potentiation (LTP) —the electrophysiological correlate of memory—at single spines while imaging Ras activity using two-photon fluorescence lifetime imaging. Ca^{2+}-dependent Ras activation spread over ~10 micrometers of dendritic length and invaded nearby spines by diffusion. Neighboring synapses along a short stretch of dendrite may thus be co-regulated due to this spread of signals downstream.

Autophagy from Egg to Embryo



Autophagy is an intracellular bulk degradation system that is critical as a self-nourishment system in the neonatal starvation period. Atg5 is a gene critical for autophagy, and knockout mice appear almost normal at birth but die immediately thereafter. It has thus been assumed that autophagy might not be important during embryogenesis in mammals. Now **Tsukamoto *et al.*** (p. 117) demonstrate that autophagy is essential for the preimplantation development of mammalian embryos. After fertilization, maternal proteins are rapidly degraded and new proteins encoded by the zygotic genome are synthesized. In the absence of maternally- or paternally-derived Atg5, autophagy is indispensable during this egg-to-embryo transition period.

Rat Run

There is a long-standing view that the functions of the dorsal and ventral hippocampus can be dissociated with respect to spatial information processing, with the dorsal hippocampus specialized for processing spatial information and the ventral hippocampus specialized for fear conditioning or defensive responses. Now, however, **Kjelstrup *et al.*** (p. 140; see the Perspective by **Hasselmo**) show that place cells exist across the entire longitudinal axis of the hippocampus, including the ventral parts that receive little or no input from relevant sensory areas of the cortex. In an 18-meter-long recording track, place cells in the dorsal one-third of the hippocampus scaled up from field sizes of 0 to 1 meter and at the ventral tip of the hippocampus scaled sizes of 5 to 15 meters. The increase in spatial scale across the hippocampus is approximately linear. The typical home range of *Rattus norvegicus* is only about 30 to 50 meters, and so the gradual increase in spatial scale is probably sufficient for simultaneous high- and low-resolution representation of the rat's entire spatial environment.

Myosins Under Tension

Myosin I is a single-headed myosin molecule that plays a role in regulating membrane dynamics and structure in eukaryotic cells. Its best-characterized function is to provide tension to sensitize mechanosensitive ion channels responsible for hearing. Myosin I is thought to function by sensing tension and changing its motile properties in response to changes in loads. **Laakso *et al.*** (p. 133) used single-molecule measurements to characterize the motor activity of myosin I. Small, resisting loads (< 2 piconewtons) resulted in a 75 times lower rate of myosin I detachment from actin, dramatically changing its motor properties. This acute sensitivity supports models in which myosin I functions as a molecular force sensor.

CREDIT: TSUKAMOTO *ET AL.*

Published by AAAS

Bruce Alberts is the Editor-in-Chief of *Science*.

Brooks Hanson is Deputy Editor for physical sciences at *Science*.

Katrina L. Kelner is Deputy Editor for life sciences at *Science*.

Reviewing Peer Review

PEER REVIEW, IN WHICH EXPERTS IN THE FIELD SCRUTINIZE AND CRITIQUE scientific results prior to publication, is fundamental to scientific progress, and the achievements of science in the last century are an endorsement of its value. Peer review influences more than just science. The Intergovernmental Panel on Climate Change and other similar advisory groups base their judgments on peer-reviewed literature, and this is part of their success. Many legal decisions and regulations also depend on peer-reviewed science. Thus, thorough, expert review of research results—without compensation—is an obligation that scientists shoulder for both science and the general public.

Despite its successes, peer review attracts its share of criticism. Reviewers can exhibit bias or only support expected, pedestrian results. They can be overtaxed, uninformed, or ask for unnecessary experiments (see Letter by Raff *et al.*, p. 36). Recent studies have explored the value of open review, double-blind review, or whether reviews are useful at all. At *Science*, we read thousands of reviews and author responses each year. From this vantage point, the system does not appear to be irretrievably broken and continues to serve science well. Reviews improve most papers, some dramatically so. Our authors sometimes thank reviewers for catching an embarrassing conclusion or for revealing a new one. We've seen peer review expose fraud (alas, not always), clarify results, and spur new insights.

But peer review is under increasing stress, in ways that are perhaps not fully appreciated. The growth of scientific publishing is placing a burden on the entire scientific enterprise. Papers today are more interdisciplinary, use more techniques, and have more authors. Many have large volumes of data and supplementary material. To compound the problem, papers are often being reviewed multiple times. Most of those rejected by *Science* go on to be considered at other journals, where the rejection rates have also increased. Before finding a proper venue, a paper may have received four, six, or even eight reviews. So even if the journal that finally publishes the article responds rapidly, the process is often painful and prolonged.

The responsibility for addressing this growing inefficiency is shared. Scientists can help by selecting the appropriate journal for their work, and seeking critical input from colleagues and all coauthors, before submitting an article for publication. Senior scientists should also mentor their students and postdoctoral fellows in good reviewing practices, enlarging the pool of qualified referees.* The possibility of repurposing reviews among journals, already practiced by some groups of journals with a single publisher, should be considered seriously. We note a recent experiment in which some independent neuroscience journals have agreed to share reviews.

The way scientists and research institutions are evaluated also needs revision. An inappropriately high value is placed on publication in certain journals. Increased competition for the limited slots in these preferred journals exacerbates the natural aggravations of peer review experienced by authors. Efforts like the Faculty of 1000, where experts scan a large set of biology journals and select the best contributions wherever published, can be very helpful. Such efforts can reduce the pressures that many group leaders feel from young scientists, who often place undue emphasis on publishing in a few high-profile journals—where the criteria used for evaluation may not match their research, no matter how valuable the contribution.

Finally, and perhaps most important, authors, reviewers, and journal editors should keep in mind the ultimate goal of scholarly scientific publishing to advance our understanding of the natural world. Competition among labs and personal striving for excellence are forces that can be harnessed to accelerate our progress. But in excess these factors can be impediments. The scientific community must collectively ensure that the peer review process continues to serve the loftier goals of our enterprise, which ultimately benefits us all.

– Bruce Alberts, Brooks Hanson, Katrina L. Kelner

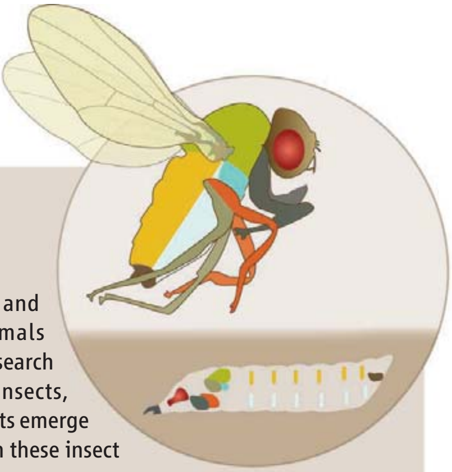
10.1126/science.1162115

**Science*'s guidelines and additional resources are available at www.sciencemag.org/about/authors/review.dtl.

CREDIT: GETTY IMAGES

Published by AAAS

EDITED BY GILBERT CHIN AND JAKE YESTON



DEVELOPMENT

Coordinated Growth

Animal species display specific developmental stages and growth rates, with individual organs and whole animals attaining a characteristic shape and size. Considerable research on growth has been performed with holometabolous insects, such as the tobacco hornworm *Manduca sexta*, where adults emerge at a size determined by the end of the last larval stage. In these insect larvae, imaginal discs specify subsequent adult organs.

Using x-ray irradiation of *Drosophila* larvae, Stieper *et al.* examine size as it is regulated by imaginal disc growth. With a low dose of x-rays, the time to final pupariation increases but ultimate size is not affected; therefore, imaginal discs adjust metamorphosis time. In addition, critical size—the minimum size of larvae at which starvation does not delay metamorphosis—increases and pupariation is delayed when ribosomal protein S3 is disrupted by RNA interference methods. — BAP

Dev. Biol. 10.1016/j.ydbio.2008.05.556 (2008).

CELL BIOLOGY

Like Ps in a Pod

In the budding yeast, the 26*S* proteasome degrades many proteins involved in cell-cycle progression and thus is essential for cell proliferation. In actively growing yeast, 80% of the 26*S* proteasome, which comprises a *20S* core particle and a 19*S* regulatory particle, is localized inside the nucleus. In quiescent cells, proteasome proteolytic activity decreases and correlates with release of the regulatory particle, but the fate of the disassembled subcomplexes remains unclear.

Laporte *et al.* found that when cells exhausted their carbon source and entered quiescence, subunits from the 20*S* and 19*S* particles colocalized into cytoplasmic foci termed proteasome storage granules (PSGs). Consistent with the proposal that PSGs act as storage depots, refeeding the cells resulted in rapid relocalization of proteasomes into the nucleus and did not require de novo protein synthesis. Other macromolecular assemblies triggered by quiescence have been described, such as P-bodies, which contain RNA and RNA-modifying proteins, suggesting that there may be a major reorganization of cellular structures upon entry into quiescence. — VV

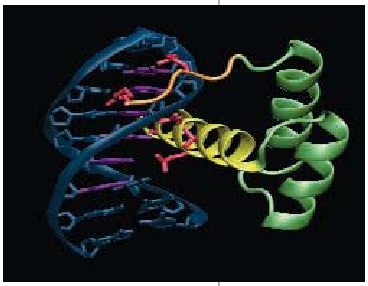
J. Cell Biol. **181**, 737 (2008).

MOLECULAR BIOLOGY

Reengineering Engrailed

The potential for engineering transcription factors so that they bind to specified DNA sequences and work as chemotherapeutic agents or sensors has generated a great deal of excitement. AT-rich sequences have been particularly challenging targets for zinc finger–domain approaches.

Noyes *et al.* have turned to the other large category of sequence-specific transcription factors, homeodomain proteins. They have carried out a comprehensive survey of the breadth of specificity of the 84 known *Drosophila* homeodomains that function independently of other DNA-binding domains. The relations between particular amino acid residues and preferred binding sequence were complex, but general determinants were assigned according to whether they cooperated or competed in binding character, leading to predictions for the binding specificity of roughly 75% of the homeodomains in the human genome and allowing them to modify Engrailed to exhibit a binding specificity resembling that of TGIF even though these proteins share only 25% amino acid identity. On the basis of their analysis, the authors have created a Web-based tool that supports the prediction of specificities for homeodomains from other organisms. — BJ



A close-up of homeodomain-DNA interaction.

Cell **133**, 1277 (2008).

MATERIALS SCIENCE

Slippery When Wet

Diamond has low friction and wear, particularly in humid environments, but the cause of this behavior is an issue of debate. One idea is that the bonds rehybridize to an ordered sp^2 form, which is consistent with graphite being the thermodynamically stable allotrope at room temperature and pressure; graphite is also an excellent lubricant because of its layered structure. An alternative idea is that the surface becomes passivated, which is consistent with data that show lower wear and friction for diamond in hydrous or H_2 atmospheres compared to experiments in vacuum. To explore this question, Konicek *et al.* created films of ultrananocrystalline diamond (UNCD), which has an extremely smooth surface and shares many of the properties of large-grained or single-crystal diamond films. Spheres coated with UNCD were rubbed against the films, either at high or low loading and high or low humidity, and the wear tracks were measured and compared with the unworn areas. The most significant wear damage occurred under high loading/low humidity conditions, which also exhibited an initially higher friction coefficient (though all four systems showed similar steady-state values). A number of techniques failed to reveal the pres-

CREDITS (TOP TO BOTTOM): (ILLUSTRATION) N. KEVITIYAGALA/*SCIENCE*; NOYES *ET AL.*, *CELL* 133, 1277 (2008)

ence of graphitic bonding, indicating that rehybridization effects were negligible and that it is rapid passivation of dangling bonds that is responsible for the low friction and wear of diamond. — MSL

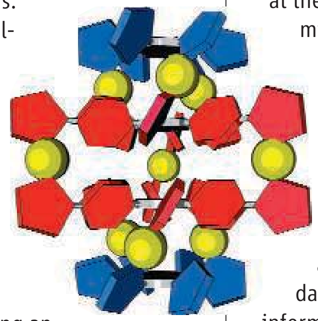
Phys. Rev. Lett. **100**, 235502 (2008).

CHEMISTRY

Spinning in Concert

In macroscopic machines, gears are commonly used to induce the synchronous motion of well-separated components. Hiraoka *et al.* observe a similar effect at the nanoscale in a stack of four ligands held together by mutual coordination to metal ions. The ligands consist of multiple oxazoline or thiazole rings appended to a central phenyl core. Upon binding silver or mercuric ions, these pendant rings adopt a common cant (shown at right) that creates an overall helicity, with the central ligands transmitting an orientational bias from one capping ligand to the other. Using solution-phase nuclear magnetic resonance spectroscopy, the authors demonstrate that a helix inversion in one component of the stack induces a cascade of inversions throughout, thereby correlating the motion of molecules spaced more than a nanometer apart. — JSY

J. Am. Chem. Soc. **130**, 10.1021/ja8014583 (2008).



PHYSICS

Quantum Privacy

For those that have it and for those that seek it, the saying that information is power is as true today as it ever was. Closely coupled to that, however, is the question of privacy—how to ensure that the information stored in a database is secure (data privacy), and that the information retrieved by users, as in a Web search, is not used against them (user privacy). For good reason, holders of information do not wish to compromise their advantage and so make it difficult to access the information (storing log files). That, however, tends to put users at the disadvantage of having to compromise their privacy or trust the database provider not to use the information in any dishonest way.

Giovannetti *et al.* show theoretically how quantum mechanics may be able to help ensure privacy for both parties. They have produced a quantum protocol that allows users to access information from a classical database without revealing which item of information it was they retrieved, and also allows perfect data privacy of the database. By quantum mechanically entangling the questions (queries would be addressed as a pulse of entangled photons, for example), any attempt by the database handler to identify which piece of information was retrieved would be scuppered as the user would be alerted. With such a quantum protocol, all parties retain their privacy. — ISO

Phys. Rev. Lett. **100**, 230502 (2008).

Science Signaling

AAAS

<< Counting Phosphates

The kinases Rad53 and Dun1 are important components of a checkpoint kinase cascade activated in response to DNA damage in yeast. Both enzymes contain forkhead-associated (FHA) domains, which bind to phosphothreonine (pThr) residues. Rad53 has four Thr residues clustered in its N-terminal SCD1 domain. Upon phosphorylation by upstream kinases, Rad53 interacts with Dun1 through the Dun1-FHA domain to activate Dun1. Although mutant Rad53 proteins that contain only one of the four SCD1 Thr residues are readily activated by upstream kinases, they cannot activate Dun1. Lee *et al*. found that a recombinant Dun1-FHA domain bound with greater affinity to Rad53-SCD1–derived phosphopeptides containing both pThr^5 and pThr^8 than to phosphopeptides that had only one of these residues, consistent with the Dun1-FHA domain, unlike that of Rad53, having not one, but two high-affinity pThr-binding sites. Treatment of yeast strains expressing a mutant *rad53* allele with a DNA-damage–inducing agent showed that the presence of both Thr^5 and Thr^8 in the Rad53-SCD1 was required for optimal Dun1 activity. Mass spectrometry studies showed the presence of monophosphorylated and diphosphorylated Rad53 proteins in response to DNA damage in vivo. Together these data suggest that whereas monophosphorylation activates Rad53, diphosphorylation of Rad53 is required to activate the Dun1-dependent arm of the DNA-damage response. — JFF

Mol. Cell **30**, 767 (2008).

CREDIT: HIRAOKA ET AL., J. AM. CHEM. SOC. 130, 10.1021/JA8014583 (2008)

Published by AAAS

www.sciencemag.org

Science

1200 New York Avenue, NW
Washington, DC 20005
Editorial: 202-326-6550, FAX 202-289-7562
News: 202-326-6581, FAX 202-371-9227
Bateman House, 82-88 Hills Road
Cambridge, UK CB2 1LQ
+44 (0) 1223 326500, FAX +44 (0) 1223 326501

Subscription Services For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

Institutional Site Licenses please call 202-326-6755 for any questions or information

Reprints: Author Inquiries 800-635-7181
Commercial Inquiries 803-359-4578

Permissions 202-326-7074, FAX 202-682-0816

Member Benefits AAAS/Barnes&Noble.com bookstore www.aaas.org/bn; AAAS Online Store http://www.apisource.com/aaas/ code MKB6; AAAS Travels: Betchart Expeditions 800-252-4910; Apple Store www.apple/eppstore/aaas; Bank of America MasterCard 1-800-833-6262 priority code FAA3YU; Cold Spring Harbor Laboratory Press Publications www.cshlpress.com/affiliates/aaas.htm; GEICO Auto Insurance www.geico.com/landingpage/go51.htm?logo=17624; Hertz 800-654-2200 CDP#343457; Office Depot https://bsd.officedepot.com/portalLogin.do; Seabury & Smith Life Insurance 800-424-9883; Subaru VIP Program 202-326-6417; VIP Moving Services http://www.vipmayflower.com/domestic/index.html; Other Benefits: AAAS Member Services 202-326-6417 or www.aaasmember.org.

science_editors@aaas.org	(for general editorial queries)
science_letters@aaas.org	(for queries about letters)
science_reviews@aaas.org	(for returning manuscript reviews)
science_bookrevs@aaas.org	(for book review queries)

Published by the American Association for the Advancement of Science (AAAS), *Science* serves its readers as a forum for the presentation and discussion of important issues related to the advancement of science, including the presentation of minority or conflicting points of view, rather than by publishing only material on which a consensus has been reached. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

AAAS was founded in 1848 and incorporated in 1874. Its mission is to advance science, engineering, and innovation throughout the world for the benefit of all people. The goals of the association are to: enhance communication among scientists, engineers, and the public; promote and defend the integrity of science and its use; strengthen support for the science and technology enterprise; provide a voice for science on societal issues; promote the responsible use of science in public policy; strengthen and diversify the science and technology workforce; foster education in science and technology for everyone; increase public engagement with science and technology; and advance international cooperation in science.

Information for Authors
See pages 634 and 635 of the 1 February 2008 issue or access www.sciencemag.org/about/authors

EDITOR-IN-CHIEF **Bruce Alberts**
EXECUTIVE EDITOR **Monica M. Bradford**
DEPUTY EDITORS **R. Brooks Hanson, Barbara R. Jasny, Katrina L. Kelner**
NEWS EDITOR **Colin Norman**

Editorial supervisory senior editor Phillip D. Szuromi; **senior editor/perspectives** Lisa D. Chong; **senior editors** Gilbert J. Chin, Pamela J. Hines, Paula A. Kiberstis (Boston), Marc S. Lavine (Toronto), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Valda Vinson; **associate editors** Jake S. Yeston, Laura M. Zahn; **online editor** Stewart Wills; **associate online editors** Robert Frederick, Tara S. Marathe; **web content developer** Martyn Green; **book review editor** Sherman J. Suter; **associate letters editor** Jennifer Sills; **editorial manager** Cara Tate; **senior copy editors** Jeffrey E. Cook, Cynthia Howe, Harry Jach, Barbara P. Ordway, Trista Wagoner; **copy editors** Chris Filiatreau, Lauren Kmec, Peter Mooreside; **editorial coordinators** Carolyn Kyle, Beverly Shields; **publications assistants** Ramatoulaye Diop, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Scott Miller, Jerry Richardson, Jennifer A. Seibert, Brian White, Anita Wynn; **editorial assistants** Carlos L. Durham, Emily Guise, Patricia M. Moore; **executive assistant** Sylvia S. Kihara; **administrative support** Maryrose Madrid

News deputy news editors Robert Coontz, Eliot Marshall, Jeffrey Mervis, Leslie Roberts; **contributing editors** Elizabeth Culotta, Polly Shulman; **news writers** Yudhijit Bhattacharjee, Adrian Cho, Jennifer Couzin, David Grimm, Constance Holden, Jocelyn Kaiser, Richard A. Kerr, Eli Kintisch, Andrew Lawler (New England), Greg Miller, Elizabeth Pennisi, Robert F. Service (Pacific NW), Erik Stokstad; **intern** Elsa Youngsteadt; **contributing correspondents** Jon Cohen (San Diego, CA), Daniel Ferber, Ann Gibbons, Mitch Leslie, Charles C. Mann, Virginia Morell, Evelyn Strauss, Gary Taubes; **copy editors** Linda B. Felaco, Melvin Gatling; **administrative support** Scherraine Mack, Fannie Groom; **bureaus** New England: 207-549-7755, San Diego, CA: 760-942-3252, FAX 760-942-4979, Pacific Northwest: 503-963-1940

Production director James Landry; **senior manager** Wendy K. Shank; **assistant manager** Rebecca Doshi; **senior specialist** Chris Redwood; **specialist** Steve Forrester; **Preflight director** David M. Tompkins; **manager** Marcus Spiegler; **specialist** Jessie Mudjitaba

Art director Kelly Buckheit Krause; **associate art director** Aaron Morales; **illustrators** Chris Bickel, Katharine Sutliff; **senior art associates** Holly Bishop, Laura Creveling, Preston Huey, Nayomi Kevitiyagala; **associate** Jessica Newfield; **photo editor** Leslie Blizard

***Science* International**

Europe (science@science-int.co.uk) **editorial: international managing editor** Andrew M. Sugden; **senior editor/perspectives** Julia Fahrenkamp-Uppenbrink; **senior editors** Caroline Ash, Stella M. Hurtley, Ian S. Osborne, Stephen J. Simpson, Peter Stern; **editorial support** Deborah Dennison, Rachel Roberts, Alice Whaley; **administrative support** John Cannell, Janet Clements, Louise Smith; **news: europe news editor** John Travis; **deputy news editor** Daniel Clery; **contributing correspondents** Michael Balter (Paris), John Bohannon (Vienna), Martin Enserink (Amsterdam and Paris), Gretchen Vogel (Berlin); **intern** Lauren Cahoon

Asia Japan Office: Asca Corporation, Eiko Ishioka, Fusako Tamura, 1-8-13, Hirano-cho, Chuo-ku, Osaka-shi, Osaka, 541-0046 Japan; +81 (0) 6 6202 6272, FAX +81 (0) 6 6202 6271; asca@os.gulf.or.jp; **asia news editor** Richard Stone (Beijing: rstone@aaas.org); **contributing correspondents** Dennis Normile (Japan: +81 (0) 3 3391 0630, FAX 81 (0) 3 5936 3531; dnormile@gol.com); Hao Xin (China: + 86 (0) 10 6307 4439 or 6307 3676, FAX +86 (0) 10 6307 4358; cindyhao@gmail.com); Pallava Bagla (South Asia: +91 (0) 11 2271 2896; pbagla@vsnl.com)

Africa Robert Koenig (contributing correspondent, rob.koenig@gmail.com)

EXECUTIVE PUBLISHER **Alan I. Leshner**
PUBLISHER **Beth Rosner**

Fulfillment Systems and Operations (membership@aaas.org); **director** Waylon Butler; **customer service supervisor** Pat Butler; **specialists** Laurie Baker, Latoya Casteel, LaVonda Crawford, Vicki Linton; **data entry supervisor** Cynthia Johnson; **specialist** Tarrika Hill

Business Operations and Administration director Deborah Rivera-Wienhold; **assistant director, business operations** Randy Yi; **senior financial analysts** Michael LoBue, Jessica Tierney; **financial analysts** Benjamin Aronin, Priti Pamnani; **rights and permissions: administrator** Emilie David; **associate** Elizabeth Sandler; **marketing director** John Meyers; **marketing managers** Allison Pritchard, Darryl Walter; **marketing associates** Aimee Aponte, Alison Chandler, Mary Ellen Crowley, Marcia Leach, Julianne Wielga, Wendy Wise; **international marketing manager** Wendy Sturley; **marketing executive** Jennifer Reeves; **marketing/member services executive** Linda Rusk; **site license sales director** Tom Ryan; **sales manager** Russ Edra; **sales and customer service** Mehan Dossani, Iquo Edim, Kiki Forsythe, Catherine Holland, Phillip Smith, Philip Tsolakidis; **electronic media: manager** Lizabeth Harman; **project manager** Trista Snyder; **assistant manager** Lisa Stanford; **senior production specialists** Christopher Coleman, Walter Jones; **production specialists** Nichele Johnston, Kimberly Oster

Advertising director worldwide ad sales Bill Moran

Product (science_advertising@aaas.org); **midwest** Rick Bongiovanni: 330-405-7080, FAX 330-405-7081; **west coast/w. canada** Teola Young: 650-964-2266; **east coast/e. canada** Christopher Breslin: 443-512-0330, FAX 443-512-0331; **uk/europe/asia** Michelle Field: +44 (0) 1223 326524, FAX +44 (0) 1223 326532; **japan** Mashy Yoshikawa: +81 (0) 33235 5961, FAX +81 (0) 33235 5852; **senior traffic associate** Deiandra Simms

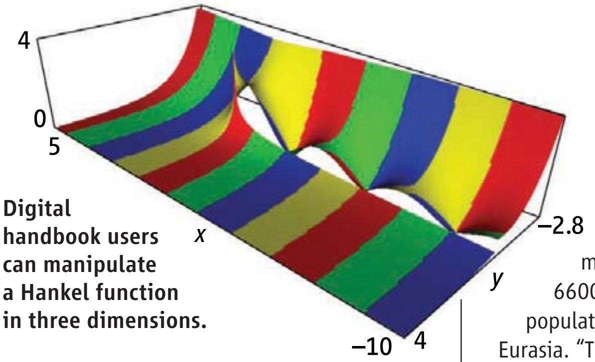
Commercial Editor Sean Sanders: 202-326-6430

Classified (advertise@sciencecareers.org); **us: recruitment sales manager** Ian King: 202-326-6528, FAX 202-289-6742; **inside sales manager: midwest/canada** Daryl Anderson: 202-326-6543; **inside sales representative** Karen Foote: 202-326-6740; **key account manager** Joribah Able; **northeast** Alexis Fleming: 202-326-6578; **southeast** Tina Burks: 202-326-6577; **west** Nicholas Hintibidze: 202-326-6533; **sales coordinators** Erika Foard, Rohan Edmonson, Shirley Young; **international: sales manager** Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532; **sales** Marium Hudda, Alex Palmer, Alessandra Sorgente; **sales assistant** Louise Moore; **japan** Mashy Yoshikawa: +81 (0)3 3235 5961, FAX +81 (0)3 3235 5852; **advertising production operations manager** Deborah Tompkins; **senior production specialists** Robert Buck, Amy Hardcastle; **senior traffic associate** Christine Hall; **publications assistant** Mary Lagnaoui

AAAS Board of Directors retiring president, chair David Baltimore; **president** James J. McCarthy; **president-elect** Peter C. Agre; **treasurer** David E. Shaw; **chief executive officer** Alan I. Leshner; **board** Lynn W. Enquist, Susan M. Fitzpatrick, Alice Gast, Linda P. B. Katehi, Nancy Knowlton, Cherry A. Murray, Thomas D. Pollard, Thomas A. Woolsey

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

EDITED BY CONSTANCE HOLDEN



Digital handbook users can manipulate a Hankel function in three dimensions.

Mathematicians' Friend Gets Blown to Bits

Every mathematician of a certain age has a shelf sporting a 6.5-centimeter-thick book: the *Handbook of Mathematical Functions*, first published in 1964 by the National Bureau of Standards, now known as the National Institute of Standards and Technology (NIST). The *Handbook*'s 1046 pages of formulas, graphs, and mathematical tables make it the definitive reference work, NIST says, on applied math's "special functions"—functions that occur frequently in modeling physical phenomena, from atomic physics to optics and water waves. Looking for the first 10 digits of Dawson's integral at 1.68, for example? It's right there on page 319. (If you need the definition, it's on page 298.)

While the slide-rule generation may cling to the printed original, NIST is getting set to unveil an online makeover of the *Handbook*, now called the Digital Library of Mathematical Functions. A beta version of five "chapters" is available at dlmf.nist.gov. The full 38-chapter library is scheduled for release early next year. For the fuddy-duddy demographic, NIST promises another 1000-page doorstop print edition.

Civilization's Tolls

Biologist Jared Diamond has hypothesized that the pathogens for diseases such as influenza and smallpox spread to humans from domesticated animals. Now, Australian and Japanese scientists say an important group of plant viruses got a boost from the spread of agriculture, which gave them a foothold by packing host plants close together.

The pathogens in question are potyviruses, which account for 15% of known plant viruses, including sugarcane mosaic virus (see photo).

Sugarcane mosaic virus.

Researchers led by Canberra virologist Adrian Gibbs compared RNA sequences of 60 species, from wild and domesticated plants around the world, to work out viral family trees. They found that the first major radiation took place about 6600 years ago, when early farming populations were spreading through Eurasia. "This modern plague only started about 200 human generations ago," says Gibbs, whose report appears online 25 June in *PLoS One*. Some potyviruses hit New World squash and papaya about 500 years ago, and others turned up in Australia after colonists arrived there 220 years ago, although potyviruses were already endemic to the continents.

The study is "very significant for world history," says Peter Bellwood, an archaeologist at the Australian National University in Canberra who has collaborated with Diamond. "This evidence supports Jared Diamond's view that human viruses underwent similar proliferation at this time."

To Sleep, Perchance to Build a New Image

"Because dreams need doing."

Can that slogan inspire a new generation of engineers? A report commissioned by the U.S. National Science Foundation says engineering needs "rebranding," and surveys show that's one of the tag lines young people like.

Messages that "students must have an aptitude for and strong interest in [math and science] to succeed in engineering" are the wrong way to go, the report from the National Academy of Engineering concludes. Instead, the authors, headed by engineering dean Don Giddens of the Georgia Institute of Technology in Atlanta, call for a "nationwide engineering awareness campaign" that depicts the field as "inherently creative, ... concerned with human welfare, ... [and] emotionally satisfying." *Changing the Conversation: Messages for Improving Public Understanding of Engineering* was released last week.

Redwoods of the Reef

For years, scuba divers have called the giant barrel sponges (*Xestospongia muta*) that sprout on the Caribbean coral reefs "redwoods" for their size and presumed old age.

A project dating the beer keg–shaped sponges now bears out that nickname. After analyzing their growth rates over 4½ years, marine biologists from the University of North Carolina (UNC), Wilmington, estimate that sponges more than a meter wide are at least 100 years old, and those larger than 2.5 meters are more than 2000 years old. One sponge, discovered off the island of Curaçao in 1997, died 3 years later. But Steven McMurray, a graduate student in Joseph Pawlik's lab at UNC and lead author of the study, calculated its age from a photograph at 2300 years. Such vintages put the sponges on a par with the oldest known California redwood, the researchers report in the current online issue of *Marine Biology*. And it makes them the longest lived animal species extant today.

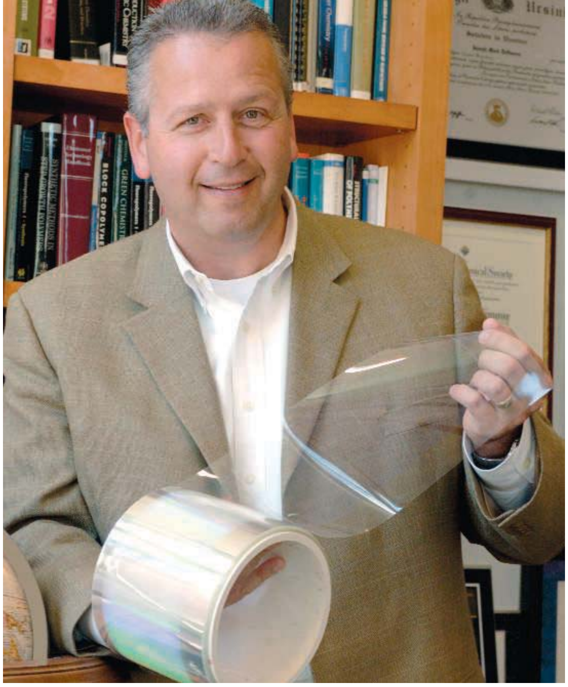
"I've seen the same individual barrel sponges on the Florida reefs for decades," says Niels Lindquist, a marine biologist at UNC Chapel Hill, so advanced ages seem "very likely." He would, however, like to see the researchers devise another way to date the sponges. Current techniques such as radiocarbon dating aren't up to the job.

CREDITS (TOP TO BOTTOM): NIST; J. R. PAWLIK/UNC WILMINGTON; UF/IFAS

Published by AAAS

NEWSMAKERS

EDITED BY KELLI WHITLOCK BURTON



<< Awards

GREENER PLASTICS. A leader in the field of green chemistry has won the 2008 Lemelson-MIT Prize. Joseph DeSimone, a polymer scientist at the University of North Carolina, Chapel Hill, and at North Carolina State University in Raleigh, received the $500,000 award 26 June during EurekaFest, an annual affair in Boston that showcases inventions by students and scientists.

DeSimone, 44, developed a technique that substitutes carbon dioxide for perfluorooctanoic acid, which can cause environmental and health problems, to make a more durable and environmentally friendly plastic called fluoropolymer. DuPont built a $40 million plant in Fayetteville, North Carolina, to produce the new material, which is used to make wire, cable insulation, tubing, and other materials for the telecommunications and automotive industries. The technology is also being tested for use in medical devices.

"The opportunity for new materials is ubiquitous," says DeSimone. "And polymer science can be a bridge to a lot of different fields to create new materials."

DEATHS

COUNTING CARNIVORES. Wildlife biologist David Maehr dedicated his career to large carnivores. On 20 June, he lost his life in a plane crash while studying them in Florida. He was 52.

Maehr made his mark in the late 1980s at the Florida Game and Fresh Water Fish Commission studying the state's endangered panthers. His research on the animals' range affected land development in the region. Although some of his findings were contested by other researchers (*Science*, 19 August 2005, p. 1162), his book *The Florida Panther: Life and Death of a Vanishing Carnivore* remains a classic, says ecologist John Gittleman of the University of Georgia, Athens.

After joining the University of Kentucky, Lexington, in 1997, Maehr turned his focus to black bears. He was conducting an aerial survey of bears near the Archbold Biological Station in Lake Placid, Florida, when the single-engine Piper Cub apparently stalled and crashed, killing Maehr and the pilot. Such research requires scientists "to hang in there for long periods of time," says Gittleman. "Dave was willing to do that."

MONEY MATTERS

RETURN TO FUNDER. Any scientist will tell you that there's never enough money for research. So NASA officials were pleasantly surprised earlier this month when a contractor saved the agency nearly $3 million.

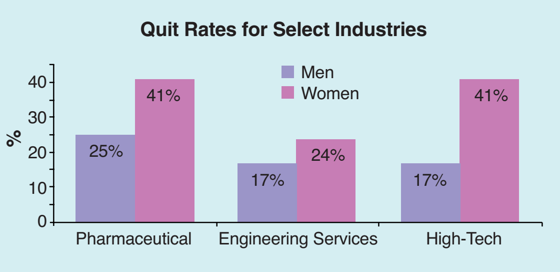
NASA budgeted $100 million in 1999 for the Solar Radiation and Climate Experiment project, a small satellite mission launched in 2003 that studies how solar radiation affects climate. A combination of experience and good luck allowed the team at the University of Colorado, Boulder, to complete the project with money to spare. Returning the funds seemed like the right thing to do, says principal investigator Thomas Woods. "NASA and the government in general are having budget problems," Woods says, so "it's good to help them out."

NASA officials say that the last time this happened was in 1996, when Johns Hopkins University's Applied Physics Laboratory shaved $3.6 million off the cost of the NEAR Shoemaker comet project.

DATA POINT

AN EXODUS OF WOMEN. A new report has found that 52% of women entering the U.S. corporate science, engineering, and technology (SET) workforce will leave their jobs at some point because of gender-related issues. One-quarter will abandon science altogether.

A survey by the Center for Work-Life Policy in New York City of 2493 men and women aged 25 to 60 found that the attrition of female SET employees spikes at about age 35 and that the pharmaceutical and technology industries are the most affected (see chart). Women reported feeling isolated—many were often the only woman in their work group—and 63% said they were sexually harassed. Other reasons for leaving included low pay and the challenges of balancing work and family.

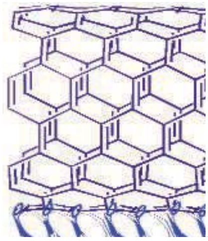


With women currently claiming 41% of all entry-level scientific jobs in industry, the impact of the hostile work environment is huge. Study authors note that reducing the number of women who quit by just one-quarter would increase the SET workforce by 220,000 people. And nearly two-thirds of the women who left SET jobs say they would gladly return if employers addressed the issues that led them to quit.

Got a tip for this page? E-mail people@aaas.org

CREDITS (TOP TO BOTTOM): COURTESY OF THE UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL; JOHN COX; NASA: SOURCE: CENTER FOR WORK-LIFE POLICY

Ancient marine diversity
24



Nanotube transistors
27

FLORIDA

Big Land Purchase Triggers Review Of Plans to Restore Everglades

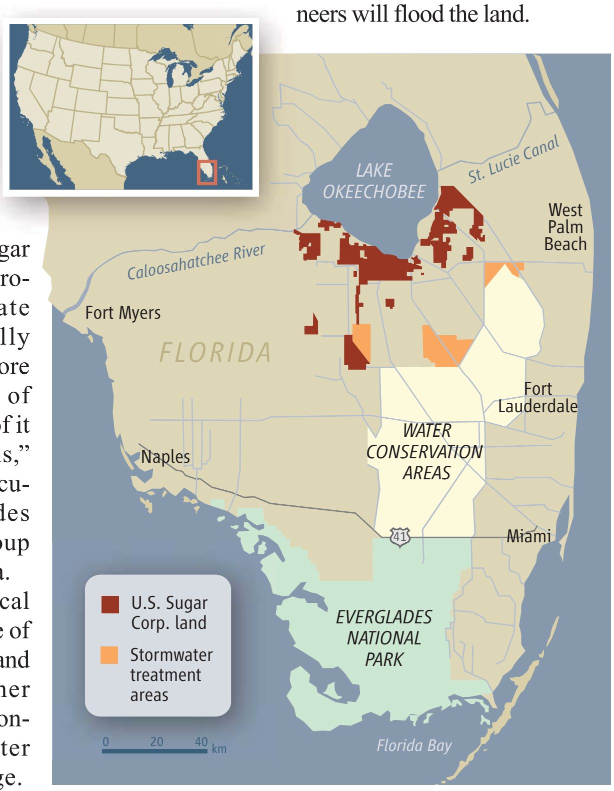
An $11 billion plan to restore the Everglades will likely get an overhaul after a major land deal last week by the state of Florida. The state's $1.75 billion purchase is intended to create wetlands that will speed up the delivery of water to the thirsty ecosystem. Environmentalists hope the two main players in the 8-year-old Comprehensive Everglades Restoration Plan—the state's South Florida Water Management District (SFWMD) and the U.S. Army Corps of Engineers—will now abandon a costly and unproven plan to store Everglades-bound water in aquifers.

Scientists say the new approach is simpler, cheaper, and big enough to accelerate recovery of the Everglades. In buying out U.S. Sugar Corp., the largest sugar cane producer in the nation, the state would recreate strategically located wetlands that could store nearly 1.2 cubic kilometers of water. "The sheer magnitude of it is beyond our wildest dreams," says Kirk Fordham, chief executive officer of the Everglades Foundation, an advocacy group based in Palmetto Bay, Florida.

Many technical and political hurdles lurk beneath the surface of the deal, however. Some of the land must be swapped with other landowners to create a large contiguous area. Improving water quality will be a big challenge. And with SFWMD spending 20% of its restoration dollars on buying land, other projects will need to be delayed or halted. The Army Corps of Engineers will also need to review its priorities, and advocates hope the agency will focus on helping the water travel south.

The Everglades needs all the water it can get. Once spanning more than 10,000 square kilometers, the wetlands were drained for agriculture and divided by roads and flood-control canals. Now less than half of the estimated 2 million cubic kilometers of water per year that used to flow south from Lake Okeechobee reaches the Everglades.

Due south of the lake, U.S. Sugar's land occupies a strategic position for restoring water flow. Under the terms of the purchase, farming will continue for 6 years while the water district negotiates land swaps with other farmers in the area to consolidate the holdings. Then engineers will flood the land.



Sweet deal. The purchase of sugar-cane fields (red) will allow Florida to assemble land for new wetlands, which will send water to the Everglades.

This means it won't be necessary, as in past wet seasons, to drain lake water seaward to prevent flooding of nearby lands. Instead, the water will be retained in new wetlands and sent south in the dry season. This approach will also reduce harm on the east and west coasts, where outflows altered the salinity of the estuaries and covered oyster beds with muck. The new storage capacity "gives us a lot more flexibility," says Tommy Stroud, the district's chief of operations.

The existing restoration plan is flawed, some say, and needs to change. It calls for water to be pumped from Lake Okeechobee into aquifers during the summer, then pumped out in the winter. Reviews by the U.S. National Academies pointed out technical problems and risks (*Science*, 9 February 2001, p. 959), and observers expect that the idea will now be ditched.

Before sending water south, the path to the Everglades must be clear. Federally run restoration projects south of the new wetlands include building channels and bridges on U.S. Route 41, which runs along an east-west levee, and filling in or reengineering canals in the Wetland Conservation Areas (WCAs). Both projects will now need to go faster, says John Ogden of Audubon of Florida. Ogden estimates that the annual $165 million federal contribution will need to be roughly doubled to get the job done.

The new plan faces other hurdles. The compacted soil in the Everglades Agricultural Area now lies a few meters below the Everglades. That means water will have to be pumped into the WCAs. Ecologist Stuart Pimm of Duke University in Durham, North Carolina, worries that the Army Corps of Engineers may go overboard in designing an engineering solution "that won't have ecological benefits for decades."

Water quality could be a problem, too. The land deal will eventually end the use of phosphorous fertilizer on U.S. Sugar's fields, which has caused eutrophication in the Everglades, and sulfate that has exacerbated mercury contamination in fish. However, both compounds remain in the soil, and still more will enter the wetlands along with water from the lake, where the sediments are highly contaminated. Aquatic plants in so-called stormwater treatment areas (see map) can remove the phosphorus but only if they are in shallow water.

Water district officials hope to finalize the deal by 30 November and begin consulting with stakeholders. Meanwhile, advocates are savoring the moment. "No one imagined we could do anything like this," says Paul Gray of Audubon of Florida. "We're in a whole new era of water management." **–ERIK STOKSTAD**

SOURCE: U.S. SUGAR CORP.

Published by AAAS

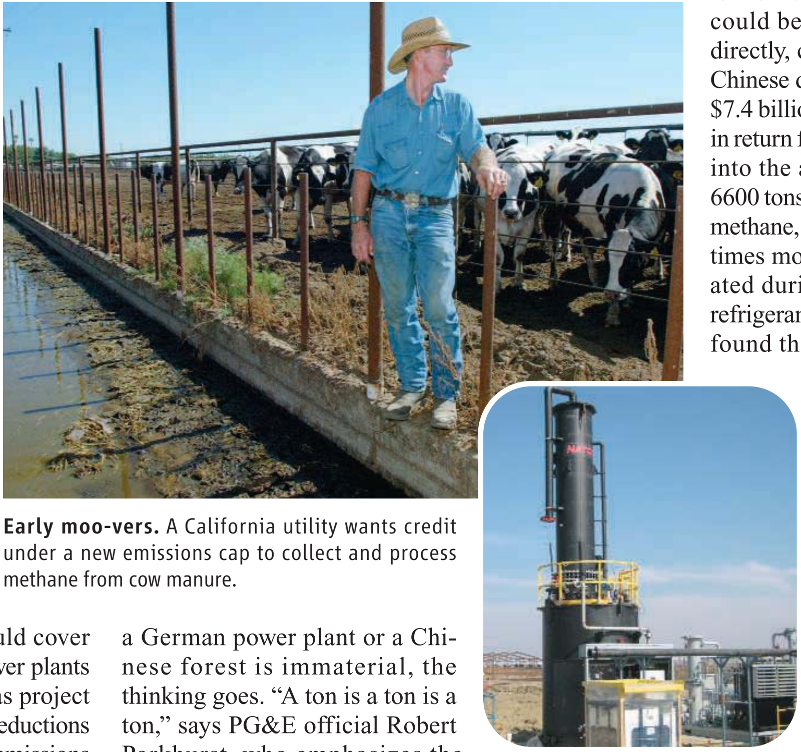
Saving Iraq's antiquities 28

Sudden death and epilepsy 31

CLIMATE CHANGE

California Emissions Plan to Explore Use of Offsets

Energy giant Pacific Gas and Electric Co. (PG&E) found a unique source of green energy last year: 5000 dairy cows on a farm near Riverdale in central California. By collecting methane from manure and turning it into natural gas for home use or electricity, officials hope to prevent the yearly emission of 1200 tons of methane, a greenhouse gas 21 times more potent than carbon dioxide (CO_2).



Early moo-vers. A California utility wants credit under a new emissions cap to collect and process methane from cow manure.

CREDIT: GARY KAZANJIAN/AP PHOTO; (INSET) PG&E

Scientists and activists applaud the company's creative effort to combat global warming. But the utility wants more than accolades. It would like bankable credit—and just might get it under rules California began drafting last week.

The plan, the most aggressive in the country, is to achieve a 10% cut in the state's current greenhouse gas emissions by 2020. A key mechanism is a system that caps the amount of emissions allowed and then permits trading in emissions credits. The system, which would begin next year, would cover everything from automobiles to power plants to factories. PG&E thinks its biogas project should offset part of the anticipated reductions that it would be required to make in emissions from its natural gas and coal facilities.

But calculating the impact of such offset projects is one of the thorniest problems facing California officials. "It's a hard question for [them]," says offsets critic Michael Wara, a former geochemist who teaches law at Stanford University in Palo Alto, California. The goal is to make sure that every dollar spent under the cap has the greatest benefit. So the challenge, he says, is to prevent companies from getting credit for "what would've happened without the incentive created by the carbon market." And how state officials deal with offsets could set a national precedent.

Offset projects are intended to encourage big emitters to reduce expected emissions cheaply and in a way that spreads the wealth. Under the Kyoto Protocol's offset program, called the Clean Development Mechanism (CDM), emitters in developed countries have purchased roughly $19 billion in credits for efforts such as forestry, agriculture, or green power projects in developing countries. Whether a company cuts emissions at a German power plant or a Chinese forest is immaterial, the thinking goes. "A ton is a ton is a ton," says PG&E official Robert Parkhurst, who emphasizes the need for a well-regulated system. "The endgame in this is reducing tons of greenhouse gas."

But experts wonder if emissions cuts claimed under CDM are really offsets or if the reductions would have happened anyway. Government incentives due to energy shortages, for example, have led to a building spree in China of low-carbon energy sources, including dams, wind power, and natural gas plants. Yet in their CDM applications, Chinese developers have claimed that Kyoto credits are the driving force behind the projects rather than pressure from the government and the expected economic payoff. Likewise, Axel Michaelowa of Germany's Hamburg Institute of International Economics has found that proposals for wind farms in India systematically left out generous government tax incentives. Stanford's David Victor estimates that up to two-thirds of emissions cuts under CDM represent cuts that would have happened anyway.

Offsets may have another flaw: They divert to inefficient projects money that could be used to cut emissions directly, critics say. For example, Chinese developers have received $7.4 billion worth of CDM credits in return for preventing the release into the atmosphere of roughly 6600 tons of a gas called trifluoromethane, a greenhouse gas 11,700 times more potent than CO_2 created during the manufacture of refrigerants. But Victor and Wara found that destroying the same amount of the gas would have cost only $157 million. This "perverse incentive," they wrote in an April paper, has fostered an industry devoted to creating the waste gas so that it can then be eliminated for cash. Worse, the extra $7.2 billion diverted to the effort could have been spent by big emitters to make real emissions cuts.

California officials, whose proposed system would cover 85% of the state's greenhouse gas emissions, say offsets could "spur innovation in unregulated sectors" such as agriculture and imported cement. They say their regulations and oversight will be stricter than CDM's, although a detailed plan won't be issued until October. In the meantime, Wara notes that California has already agreed to allow its industries to trade in emission certificates from its neighbors—some of whom are bound to accept CDM credits, tainted or not—under the 10-state Western Climate Initiative launched last year. "We are really excited about [PG&E] doing this project," says attorney Kristin Grenfell of the Natural Resources Defense Council's San Francisco office. "We just don't think that offsets are the best way of encouraging it."

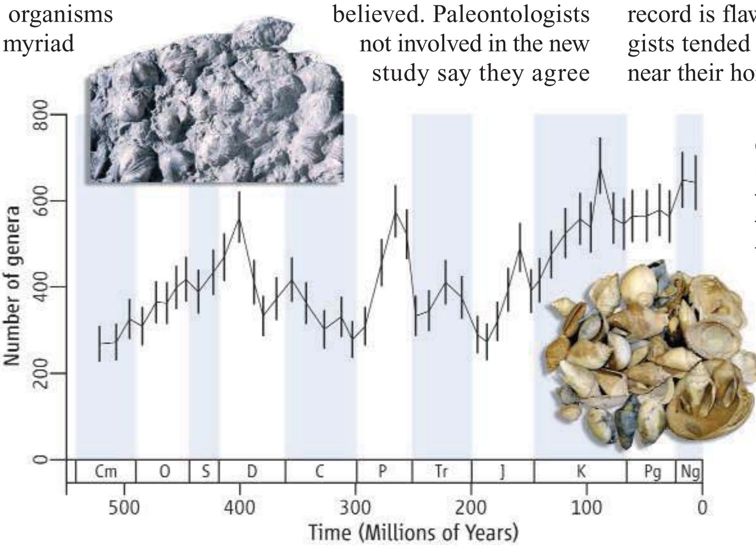
–ELI KINTISCH

Published by AAAS

PALEONTOLOGY

Life's Innovations Let It Diversify, at Least Up to a Point

When paleontologists first seriously considered how life had evolved, the story looked simple: From a few basic types, organisms had diversified unhindered into myriad new forms over the past half-billion years. Then the bean counters got into the act. After correcting for sampling biases and other pitfalls in the fossil record, a group of quantitatively oriented paleontologists reported in 2001 that life at sea, at least, had followed a different course (*Science*, 25 May 2001, p. 1481). Although marine invertebrates had continued to innovate new ways of making a living, the scientists concluded, total diversity had hardly increased in 400 million years.



Diversity constrained? The latest curve of marine diversity lacks a recent sharp rise, suggesting something reined in evolution. Alternatively, excluding young fossils found in loose sediment (near left) may have damped the rise.

Now the number crunchers have rewritten the prehistory books again. On page 97, 35 of them—including authors of the original paper—present a new analysis of the Paleobiology Database, which records about 3.5 million specimens described in papers of the past century and more. They conclude that the diversity of marine invertebrates has indeed increased over time, although far less than some early analysts believed. Paleontologists not involved in the new study say they agree with its general approach but doubt it will be the last word on the subject. "There's a lot of improvement in methodology, and there's a lot more data," says David Jablonski of the University of Chicago in Illinois, who worked on the 2001 analysis but not the new one, "but there are still biases remaining in the data that remain to be addressed."

Everyone agrees that the raw fossil record is flawed. For 180 years, paleontologists tended to collect their favorite fossils near their home institutions in North America and Europe, neglecting fossils in remote lands. They collected more and smaller fossils from young, loose sediments than they did from older rock. And of course they were more likely to collect the fossils that tended to be preserved rather the more vulnerable ones that fade away with time.

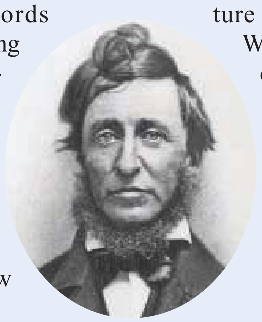
The Paleobiology Database—compiled under the supervision of John Alroy of the University of California, Santa Barbara (UCSB)—includes information needed to correct such biases in the record, information

EVOLUTION

Where Have All Thoreau's Flowers Gone?

Early in May, the oaks, hickories, maples, and other trees, just putting out amidst the pine woods around the pond, imparted a brightness like sunshine to the landscape. …

These words from *Walden* hint at the careful plant and animal records Henry David Thoreau kept during his stay at Walden Pond in Concord, Massachusetts, in the mid-1800s. By retracing this young naturalist's footsteps, not once but twice in the past century, researchers have been able to chronicle the fate of hundreds of plant species as the New England climate has changed since Thoreau's time. Using that data, Harvard University graduate student Charles Willis and colleagues have detected a disturbing pattern, one that he described last week in Minneapolis, Minnesota, at the Evolution 2008 meeting.



Flower man. Thoreau closely tracked Walden Pond plants

By building a flora family tree that incorporates the "Thoreau" species and mapping onto the tree each plant's response to the 2°C increase in the region's average temperature since the famed author was at Walden Pond, the researchers have discovered that climate change has placed whole groups of plants at risk and that the more charismatic wildflowers that prompt conservation efforts, such as orchids, are among the most vulnerable.

The study is "an intriguing combination of historical data sets and modern molecular methods to address in a very novel way climate change effects," says Carol Horvitz, a plant evolutionary ecologist at the University of Miami, Florida. "I think it's brilliant."

Many studies have looked at how global warming may cause shifts in where plants grow, but very few have examined how specific traits, such as flowering time, are affected. The necessary long-term records rarely exist. But for 6 years, Thoreau tracked the life histories of more than 400 plant species in a 67-square-kilometer area. Another researcher covered the same ground at Walden Pond and its surrounds circa 1900. Then from 2004 to 2007, Boston University (BU) conservation biologist Richard Primack and his student Abraham Miller-Rushing regularly visited the area to make similar observations of about 350 species and to check how the abundances of these plants had changed through time.

Their data, published in February in *Ecology*, revealed that many flowers were blossoming a week earlier than in Thoreau's time. They noted also that about half of the species studied had decreased in number, with 20% having disappeared entirely.

Working with his Harvard adviser Charles Davis, the BU group, and fellow Harvard graduate student Brad Ruhfel, Willis has put

CREDITS: (GRAPH) J. ALROY ET AL., *SCIENCE*; (IMAGES, TOP TO BOTTOM) D. L. MEYER/UNIVERSITY OF CINCINNATI; S. KOLBE; COURTESY OF THE THOREAU INSTITUTE AT WALDEN WOODS

such as exactly where in the record and in what sort of rock each fossil was found. Alroy, lead author of both the 2001 study and the new paper, says both groups also applied statistical techniques as they "sampled" the database to ensure that their count resembled reality. Unfortunately, he says, the 2001 team made some assumptions about sampling that "turned out to be dramatically wrong" in ways that would have made an increase in diversity through time hard to find. The new analysis corrects those errors, he says. It's also based on four times as much data spanning all of the past 500 million years.

The resulting graph of changing diversity over time resembles the pre-2001 curve in showing a steep rise in diversity in the first 100 million years. The curves differ most sharply in the past 65 million years, when diversity soars dramatically on the pre-2001 curve but hardly rises on the new one. As a result, the number of genera in geologically recent times appears to have increased only about 30% over life's early peak, not three- to fourfold as the old curve showed. Something has been constraining evolution and diversity for hundreds of millions of years, the group concludes—perhaps some bottleneck in the way energy moves up through the food chain in the global ecosystem.

Although the latest diversity curve marks a big improvement over the 2001 effort, it may go too far, says paleontologist Richard Bambach of the Smithsonian National Museum of Natural History in Washington, D.C., another co-author of the earlier paper but not of the current one. "We're getting into the ballpark, [but] they're taking the most conservative approach," he says. The newly estimated diversity of the past 10 million years in particular may be "excessively conservative," he says.

For one thing, Bambach says, the group excludes all fossils recovered from sediments that have not yet turned to stone. That makes sense in principle, he explains. Because sieving loose sediments for fossils is so much easier than breaking rocks, including fossils from silt and mud could inflate the apparent diversity of more recent times, when most such "unlithified" sediments are found. On the other hand, if diversity really has increased recently, ignoring younger samples could seriously undercount it, Bambach says. Jablonski also suspects that younger diversity is being missed in the western tropical Pacific Ocean. Today, shellfish are wildly diverse there, he notes, but in the database they appear to be relatively impoverished only a few million years ago. More likely, he says, the database has yet to include the older literature from that region. Thus, some observers are looking for a third iteration of life's changing diversity. –RICHARD A. KERR

Climate shift. The warming of the Walden Pond area (*above*) since the 1880s threatens many plants.

CREDIT: JOSEPH SOHM/VISIONS OF AMERICA/CORBIS

these data into an evolutionary context by looking at how closely related the affected species were. They pieced together a family tree of more than 500 species and noted changes in their range, abundance, and other traits—such as which had flowering times that were tied to spring temperatures and which did not. In this way, they could check to see if there was a correlation between flowering time and how well a species fared over 1.5 centuries.

"Certain [groups] were very sensitive," Willis reported at the meeting. A plant's ability to change its flowering time depending on the spring weather in a given year proved a key predictor of its current health. "Species that had not shifted [flowering times] are declining in abundance."

Wildflowers with more northern ranges proved the least flexible. Thus irises, orchids, lilies, and bladderworts were among the plants that had declined the most—they tended to flower the same time of year, regardless of the weather. "That climate change is affecting whole sets of species differently is intrinsically interesting," says Horvitz.

Rare as they may be, these sorts of analyses can help researchers predict which species are threatened most by global warming and which are likely to adapt, says George Weiblen, an evolutionary biologist at the University of Minnesota, Minneapolis. "Finally, evolutionary biologists are chiming in on the effects of climate change." –ELIZABETH PENNISI

SCIENCESCOPE

Peruse These Ties

Congressional appropriators want to give the U.S. National Institutes of Health (NIH) a billion-dollar budget increase next year. But senators also want NIH Director Elias Zerhouni to stiffen his agency's oversight of the financial ties between academic scientists and pharmaceutical companies.

Last week, the Senate spending panel ordered NIH's parent body, the Department of Health and Human Services, to begin the process of rewriting regulations meant to avoid financial conflicts of interest among nonfederal scientists. Grantees now must report to their institutions relevant income from any company exceeding $10,000 annually. (NIH intramural scientists are banned from receiving all such income.) A House spending subcommittee has asked NIH directly to improve its conflicts policy. NIH is already planning to seek comments on revising the regulations, Zerhouni said in a 20 June letter to Senator Charles Grassley (R–IA), who is investigating several cases in which academic researchers may have failed to report income from drug companies (*Science*, 27 June, p. 1708).

The Senate language is part of a 2009 spending bill that is unlikely to be approved until after the November elections. But agencies ignore such congressional requests at their peril. The bill itself would provide NIH with a $1.025 billion hike, to $30.2 billion, the agency's largest increase in 6 years. A parallel House bill would give NIH a $1.15 billion increase. –JOCELYN KAISER

Austrian Astronomers Score

VIENNA—Last week, the Austrian government joined the European Southern Observatory (ESO), Europe's premier telescope facilities, based in Chile's Atacama Desert. "At meetings, everyone always assumed we were members," says Josef Hron, an astrophysicist at the University of Vienna. Cash-strapped Austria has declined membership since ESO was established in 1962, he says, but the economy and astronomy have flourished in the country over the past decade. Considering the $3.6-million-per-year cost of membership—plus a $36 million entrance fee to be paid over 15 years—Austrian astronomers have just seen their budget doubled. Not only will they have far easier access to ESO's Very Large Telescope—currently the largest of its kind—but they plan to take an active role in future ESO projects, such as ALMA and the Extremely Large Telescope.

–JOHN BOHANNON

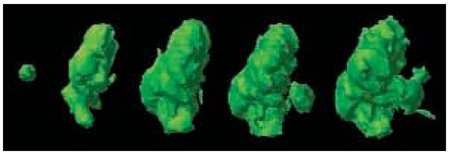
Published by AAAS

GENOMICS

Billion-Dollar Cancer Mapping Project Steps Forward

Leaders of an ambitious effort to find all common mutations in human cancers delivered their first results to a U.S. government panel last week along with a plain message: Their $100 million pilot is paying off. A sweeping search for mutations in one cancer—glioblastoma, a type of brain tumor—has verified known genes and turned up a few new ones, said lead presenter Eric Lander of the Broad Institute in Cambridge, Massachusetts. The data from DNA sequencing of tumors, combined with other genetic analyses, he said, are already pointing to potential new therapies to extend the lives of glioblastoma patients, who now rarely survive much longer than a year.

Lander and others were making the case for a massive, multiteam endeavor on the scale of the Human Genome Project, known as The Cancer Genome Atlas (TCGA). First proposed by Lander 3 years ago, TCGA would aim to find all common mutations in the major human cancers over 10 years, at a cost of up to $1.5 billion. After some scientists panned the idea as not worth the price tag (*Science*, 21 October 2005, p. 439), the National Cancer Institute (NCI) and the genome institute launched a 3-year pilot project.



Target number one. Glioblastoma, a cancer of the brain, was studied in a pilot gene-sequencing project.

The audience for last week's pitch, NCI's Board of Scientific Advisors (BSA), was listening on this occasion, not voting. But at least one member who calls himself a former skeptic—Lee Hartwell of the Fred Hutchinson Cancer Research Center in Seattle, Washington—says he was impressed: "At this early stage, to come up with something that essentially changes how we think about [glioblastoma] therapy is pretty phenomenal."

Lander summed up the search for genes involved in glioblastoma, the first of three cancer types under investigation. After sequencing 600 candidate genes in 86 samples, TCGA has verified five genes already known to be mutated in glioblastoma and found three new ones, including *NF1*, which seems quite important—this gene was inactivated in 20% of the samples. Other teams, who are combining these genes with data on gene expression and gene copy number, report that primary glioblastoma appears to occur as three or four distinct subtypes. And Charles Perou of the University of North Carolina, Chapel Hill, noted that each subtype shows different patterns of mutations within key gene signaling pathways.

These results suggest that clinicians ▶

MARINE BIOLOGY

Panel to Take Closer Look at Scientific Whaling

Is Japan's research whaling, which claims about 1000 whales a year, scientific investigation or disguised commercial whaling? A new review process endorsed by the International Whaling Commission (IWC) at its 60th annual meeting held last week in Santiago, Chile, hopes to get closer to the answer.

Departing from past practice, the review panel will likely include experts from outside IWC's Scientific Committee and exclude scientists involved in Japan's research. Even so, says John Frizell, a spokesperson for Greenpeace International, which has criticized Japan's effort, the results "will depend on who the outside experts are."

Japan carries out one hunt a year in Antarctic waters and another in the North Pacific. The Tokyo-based Institute of Cetacean Research (ICR) maintains that the animals must be killed to obtain data, such as age and stomach contents, needed to develop management plans for the resumption of commercial whaling. During 2008–09, Japan plans to take 1330 whales, primarily minke, with limited numbers of sei, Bryde's, sperm, and fin whales. The total, which includes 50 humpback whales, is more than all other countries kill for scientific, commercial, and subsistence purposes combined. But Japan has "postponed" its humpback catches pending further IWC discussions (*Science*, 27 April 2007, p. 532).

Unprotected. Japanese whalers capture short-finned pilot whales, which are not covered by a moratorium on commercial whaling.

Under IWC rules, the Scientific Committee reviews research programs before they begin and then every 6 years. In the past, Japanese scientists were part of that review team and helped write the reports. "That is not the way reviews are done in the real world," says one Scientific Committee member, speaking on condition of anonymity.

The new approach, adopted when the committee met in Santiago ahead of the annual meeting, will allow scientists conducting the research to present results and answer questions but not serve on the review panel. The process will be used to analyze Japan's ongoing North Pacific program next spring.

In its report to IWC, the Scientific Committee says the changes are intended to improve the review process. But some also see it as implicit criticism of Japan's whaling program, with one committee member noting that they hope "to submit bad science to a proper review." Dan Goodman, an adviser to ICR in Tokyo, says the institute hopes to use the results from the new review process to refine its research whaling program, which he says is "a right of every contracting government" under the IWC convention.

Meanwhile, IWC has set up a new working group to propose compromises on divisive issues such as research whaling and creating whale sanctuaries. The group will begin its work long before the Scientific Committee completes its review, however.

–DENNIS NORMILE

CREDITS (TOP TO BOTTOM): ARTHUR TOGA/UCLA/PHOTO RESEARCHERS INC.; ISSEI KATO/REUTERS/LANDOV

could classify glioblastoma patients by their tumor types, then tailor therapy to the genes or pathways that matter most, reported neurosurgeon Cameron Brennan of Memorial Sloan-Kettering Cancer Center in New York City. He cautions, however, that not all the data are in hand.

The BSA had some questions for the scientists. One member found it "unsettling" that patients with different subtypes all had the same survival rates. And chair Robert Young, president of Fox Chase Cancer Center in Philadelphia, Pennsylvania, says these early results probably won't win over some skeptics: "It's still debatable whether sequencing is the only or best way" of exploring the cancer genome, Young says. Still, "this first presentation is encouraging," Young says. "It's doable."

–JOCELYN KAISER

APPLIED PHYSICS

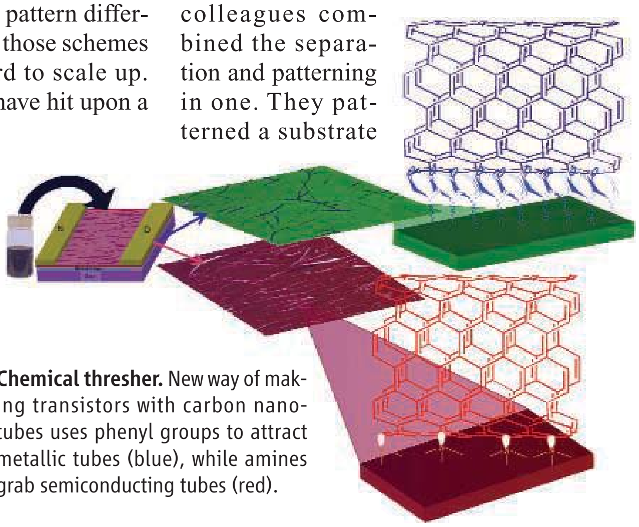
Tiny Transistor Gets a Good Sorting Out

For electronics researchers, carbon nanotubes are like opera divas—full of power and headaches. The tiny, all-carbon tubes carry oodles of electric current for their size, and they can behave as either metals or semiconductors depending on their atomic arrangement. Electronic devices such as transistors and wires are best made with either semiconducting or metallic tubes, not both—yet, when produced, carbon nanotubes come out as a mix. Researchers have devised ways to separate and pattern different types of tubes. But so far, those schemes have been complex and hard to scale up. Now a California team may have hit upon a simple solution.

On page 101, researchers at Stanford University in Palo Alto, California, led by chemical engineer Zhenan Bao report using different chemical compounds to attract metallic and semiconducting tubes to different areas on a surface. Using this approach, the researchers separated and patterned semiconducting tubes in one step to form the heart of a transistor that turned off and on much more efficiently than previous transistors made with bunches of nanotubes. "This is good work" and an "important piece" toward the overall goal of integrating carbon nanotubes into high-performance electronics, says Jeffrey Bokor, an electrical engineer at the University of California, Berkeley.

Electrical engineers have turned to carbon nanotubes and other materials in recent years as the push to make ever-smaller transistors continues to reduce the amount of current that silicon can shuttle in the critical channel between electrodes. Given their ability to carry large currents, carbon nanotubes may do better. But to make good devices, researchers must span the electrodes with tightly packed rafts of nanotubes. Those rafts must contain only semiconducting tubes, not metallic ones, which can't switch on and off.

Prior research groups have come up with schemes to lay down both types and then burn out the metallic tubes, leaving the semiconductors behind. Another technique separates the two in solution before patterning them. In their current work, Bao and her colleagues combined the separation and patterning in one. They patterned a substrate with simple compounds called amines that then attracted just the semiconducting tubes to lie down and bind. Washing the surface removed any metallic tubes and aligned the semiconducting ones in the channels between electrodes. The result was individual transistors that, unlike many previous devices, conducted far better when switched on than they did when turned off.



Chemical thresher. New way of making transistors with carbon nanotubes uses phenyl groups to attract metallic tubes (blue), while amines grab semiconducting tubes (red).

CREDIT: Z. BAO/STANFORD UNIVERSITY

The next step for Bao's group and others, Bokor says, is to pack more tubes in the channels to carry more current. If they do that, nanotube transistors may soon give silicon a run for its money.

–ROBERT F. SERVICE

SCIENCESCOPE

Settlement in Anthrax Case

Nearly 6 years after naming biomedical scientist Steven Hatfill a "person of interest" in its investigation of anthrax-laced letters, the U.S. Justice Department has agreed to pay him $5.8 million. Hatfill, who formerly worked in the Army's Fort Detrick, Maryland, biodefense laboratory, claimed in his 2003 suit that the government had invaded his privacy in its quest to solve the 2001 incidents, in which five people died. Hatfill's attorney says the settlement, for which the government admits no wrongdoing, means "justice" for his client.

–ELI KINTISCH

Rights for Apes? ¡Sí!

Spain's parliament plans to give great apes the right to "life, liberty, and freedom from torture," a move that has few practical implications in Spain but that supporters hail as a landmark that could make more countries consider adopting such measures. The environmental committee in Spain's Congress of Deputies approved the bill last week; it is expected to become law.

The bill instructs Spain's government to adhere to, and promote in the European Union, the Great Ape Project (GAP), a movement started in 1993 by scientists and philosophers that aims to grant the apes basic legal and moral rights. As a result, harmful scientific studies—which are no longer carried out in Europe—would be banned. "We really haven't seen this before for any animals at any national level," says Princeton University bioethicist and GAP co-founder Peter Singer.

–MARTIN ENSERINK

Alliance Aims for Cancer Vaccine

The GAVI Alliance, a global partnership that helps poor countries buy vaccines, plans to branch out into seven new diseases—including cervical cancer, a disease targeted by two new vaccines whose cost looms as a major obstacle to poor countries (*Science*, 16 May, p. 860). Last week, GAVI's board also decided to add cholera, typhoid, meningitis A, rabies, Japanese encephalitis, and rubella to its list of targets.

Vaccines against the human papillomavirus (HPV), which causes cervical cancer, have been widely introduced in Europe and the United States. GAVI is hoping producers will agree to lower the current price of $360 for three doses to about $21, says GAVI policy and strategy director Nina Schwalbe. "This is a milestone," says Joakim Dillner, an HPV expert at Lund University in Sweden. "GAVI is the scientific community's only hope for bringing this vaccine to developing countries."

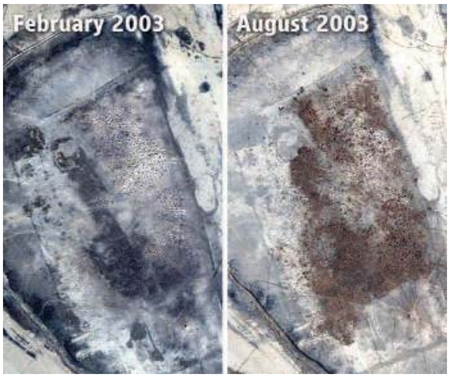
–MARTIN ENSERINK

Preserving Iraq's Battered Heritage

Archaeologists have feared for Iraq's unique archaeological treasures since war began 5 years ago. Now, despite continued unrest, a team returning from southern Iraq bears surprisingly good news

Online
sciencemag.org
Podcast interview with author Andrew Lawler

IN THE EARLY MORNING LIGHT OF 7 JUNE, an international team of archaeologists examined the ancient settlement of Tell al-Lahm in the flat and fertile plain of southern Iraq for signs of looting. Then three pickup trucks with armed men suddenly arrived on the scene. What followed was a brief but welcome confrontation: The men were part of a security team tasked with protecting such lonely sites from artifact thieves. Five years after the U.S. invasion of Iraq and the overthrow of Saddam Hussein plunged the country into chaos and sparked a looting spree, a semblance of order is returning to the home of humanity's first writing system, cities, and empires. The team—the only group of scholars allowed to visit the area since the summer of 2003—found guards patrolling several sites and little evidence of extensive theft in recent years. "It's really good news after so many years of doom and gloom," says archaeologist and team member Elizabeth Stone of Stony Brook University in New York state.



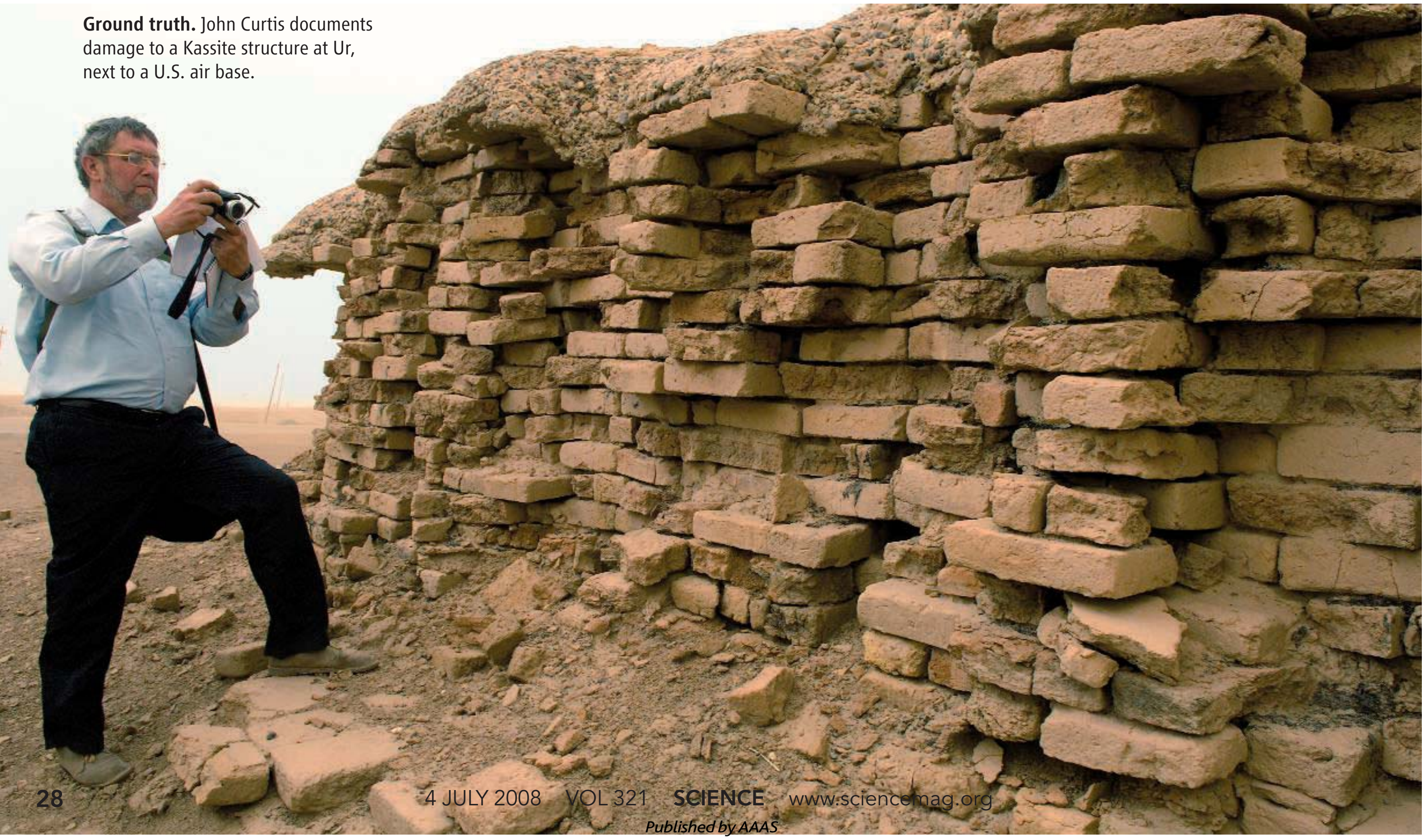
Overturned. Looters decimated the ancient Sumerian city of Kisurran in southern Iraq, as seen by turned earth (brown) in the second satellite image.

That good news extends to Amman, Damascus, and New York, where investigators succeeded recently in recovering thousands of artifacts stolen from Iraq after the invasion. European governments are also moving, albeit slowly, to help rehabilitate Iraq's shattered museums, rebuild fences at exposed sites, and provide remote-sensing data to researchers. Serious difficulties remain. Iraqi archaeologists trying to protect priceless artifacts and ancient settlements still face intimidation and even jail time (see sidebar, p. 29). Allied troops damaged some of the most famous cities of the ancient world and have yet to address the problems. And the international market in Mesopotamian antiquities continues to thrive, likely fueled by continued looting at smaller and more remote sites. But archaeologists are finally gaining access to the region, allowing them to make realistic appraisals and recommend action.

Sky view, ground truth

Before the trip, archaeologists had already seen signs of widespread looting, primarily with the help of satellite data. In a paper in

CREDITS (TOP TO BOTTOM): DIGITAL GLOBE/ELIZABETH STONE; BRITISH MINISTRY OF DEFENSE



Ground truth. John Curtis documents damage to a Kassite structure at Ur, next to a U.S. air base.

Published by AAAS

the March issue of *Antiquity*, Stone used remote-sensing images to examine 101 square kilometers of archaeological sites in southern Iraq. She found nearly 16 square kilometers of looting holes in that area—four times the amount of legal excavations conducted in the same areas during the past century. Much of the looting appears to have taken place during two periods. The first was in the mid-1990s—a time of desperate poverty in southern Iraq—and the second during the spring and summer of 2003, when allied troops were fighting Saddam's army and then the growing insurgency.

But until last month's expedition, no team of foreign archaeologists had actually been on the ground in southern Iraq since mid-2003. The group, which planned to reveal its findings at a 4 July press conference in London, included researchers from the United States, the United Kingdom, and Germany, as well as two Iraqi officials. They were sponsored by the British military, which provided a helicopter and security team. During their 3-day visit, the researchers visited and assessed eight major sites, including Eridu, home of an early temple complex, and Uruk, where legend says the world's first writing system, cuneiform, was developed. "It was not as bleak as we feared," says John Curtis, an archaeologist at the British Museum in London who took part.

The major exception was Ur, famed center of a Sumerian city-state 4 millennia ago and home to a partially reconstructed ziggurat. A large American air base sits immediately adjacent to the site, which has drawn hordes of visiting American soldiers. The result is extensive wear and tear, team members say. For example, a building from the Kassite era, circa 1400 B.C.E., is in danger of collapse. "One of the earliest arches in the world is going to fall down," says Stone. American engineers also bulldozed a nearby 2nd millennium B.C.E. site as part of a base expansion between August 2004 and August 2005, team members say.

The base's location violates a decree by Iraq's State Board of Antiquities and Heritage mandating a 500-meter-wide protective zone around archaeological sites. But U.S. State Department spokesperson William Olson in Baghdad refused to comment on any damage sustained at Ur or on whether the U.S. government intends to fund security or restoration projects there. Curtis says the British government may seek money to fix some of the most egregious damage at Ur and other southern sites.

Consistent security has helped preserve nearby Uruk, where King Gilgamesh is said to have reigned. One guard is paid by the

CREDIT: ABDEL-AMIR HAMDANI/SBA

ON IRAQ'S FRONT LINES

First Abdel-Amir Hamdani was charged with theft and kidnapping. Then his home was searched. Finally, the director of antiquities in the Nasiriyeh region of southern Iraq was thrown into jail for 3 months. His crime? Opposing plans by unscrupulous developers intent on building a dozen brick factories on top of an important archaeological site. But Hamdani, who was ultimately cleared of all charges and released, says his harrowing ordeal was worth it because the developers eventually abandoned their plans. "The result was good," he said during a recent visit to the United States. "If you gave me a choice between jail and brick factories, I would choose jail."

Hamdani's tribulations reflect the precarious state of the country's archaeological heritage 5 years after the war began (see main text). "This is what we have to do as archaeologists to protect Iraq's heritage," says Donny George, former chair of the State Board of Antiquities and Heritage in Baghdad and now a professor at Stony Brook University in New York state.

Shy, soft-spoken, and rail-thin, the 40-year-old Hamdani is an unlikely archaeological hero. But since the U.S. invasion began, he has tried to find ways to protect the vast region he oversees in south-central Iraq. Intensive looting began as soon as Saddam Hussein's forces retreated and the U.S. military rolled north to Baghdad in March 2003, he says. The region is littered with thousands of ancient settlements, which represent a treasure trove of salable goods to an impoverished population. Hamdani has worked with a succession of American and Italian military officers to ensure that archaeological sites were patrolled; he even traveled to the holy city of Najaf to explain the dire situation to Grand Ayatollah Ali al-Sistani, a powerful figure in mostly Shiite southern Iraq. Al-Sistani quickly issued a fatwa forbidding the pillaging of ancient sites. Hamdani is also currently surveying sites exposed in the past decade after Saddam drained the marshes of southern Iraq.

Setting his sights. Hamdani searches the Iraq marshes for ancient settlements.

Such work is complicated by continued fighting, however. The museum's offices were ransacked and the adjacent library burned in a May 2004 clash between militia and the Italian forces who replaced the Americans. "We lost everything," recalls Hamdani. The Italians later renovated the museum, but this spring attacks destroyed vehicles and damaged the museum's façade and roof.

Hamdani's biggest challenge was deflecting a 2005 proposal by the Nasiriyeh city council to build 12 brick factories just outside town, between the ancient Sumerian cities of Ur and Ubaid. The site of very early settlement dating to the 6th millennium B.C.E., Ubaid gives its name to an entire era known as the Ubaid period. Ur was a large city during the first florescence of urban areas in the 3rd millennium B.C.E., as well as the legendary home of Abraham. When Hamdani conducted a required survey, he discovered that the site was littered with ancient Sumerian material. "We need these factories," he says, "but not on top of an archaeological site." So, representing the Baghdad antiquities department, he denied permission for construction.

In February 2006, Hamdani says that those supporting the site location struck back with a memo to a local judge alleging that he had stolen gasoline from departmental tanks, that he was involved with a kidnapping, and that his son was an antiquities smuggler. Police searched his home and found nothing suspicious, but that April Hamdani was jailed. He calls the accusations absurd, given that the department has no gasoline tanks in Nasiriyeh and that his son was 2 years old at the time. And he denies any involvement in kidnapping. George confirms the tale and says that Hamdani's success in putting looters in jail led to the reprisal. After officials in Baghdad intervened, Hamdani was cleared of the charges and released that June.

The experience has not cowed Hamdani, who studied archaeology at Baghdad University. But he feels lucky to have escaped the ordeal with his life. "I could have been shot like so many others," he says. "There is an underworld there like the Mafia. Sometimes you forget being an archaeologist, and you work as a policeman." **–A.L.**

Published by AAAS

German government, and 15 more are part of the Iraqi Facilities Protection Service (FPS) set up in 2003 to protect Iraqi government sites. Margarete van Ess, an archaeologist at the German Archaeological Institute in Berlin and the third Western member of the team, was delighted to find her pottery and excavation materials intact in the dig house. "I guess I could go there and restart my research," she says, adding that she hopes to return once the situation is less dangerous.

Market of thieves. These looted antiquities were confiscated in a small town in southern Iraq, where business in artifacts is brisk.

Curtis, Van Ess, and Stone say they were heartened by the FPS guards who confronted them briefly at Tell al-Lahm. FPS has been accused within Iraq of becoming a militialike organization, but archaeologists say they are grateful for their presence. "It is very encouraging that the efforts to protect sites really have worked," adds Stone. But Abdel-Amir Hamdani, Iraq State Board of Antiquities inspector in the Nasiriyeh region, says the Iraq government has refused to provide fuel since 2006 for FPS patrols, hampering their effectiveness.

Spy operation

The three archaeologists agree that their limited visit provides little new data on the host of other sites in southern Iraq that satellite data suggest may still be plagued by looting. Hamdani says that smaller and more remote sites are especially vulnerable. The international team was unable to visit any of these sites, although Stone confirmed that remote-sensing images show widespread damage to ancient settlements in the area.

Lacking the firepower to take on the often-armed bands that denude sites, Hamdani has tried to recover some of their plunder and catch the ringleaders by posing as a buyer at two villages known as centers of the black market, El Fajir and Albhagir, on the northern border of Dhi Qar governate. During one visit, a village boy asked him what kind of artifacts he wanted—gold objects, tablets, statues—and led him to the appropriate dealer. Hamdani was then able to tip off Italian soldiers and Iraqi police. In a single home, they discovered 600 artifacts, including pots and cuneiform tablets. Albhagir was once a typical small and impoverished southern Iraqi village, says Hamdani. Now it boasts large homes that he suspects are funded with money from the illegal trade. "Some 70% of the population work in looting and smuggling artifacts," he estimates. "He's now running a spy operation" using informants, says Stone with admiration.

But sometimes a successful sting operation isn't enough. In November 2004, for example, a truck carrying recovered looted material on its way to the Iraq Museum in Baghdad was hijacked, the driver and guards killed, and the artifacts stolen.

Thousands of looted objects have slipped across Iraq's porous borders since 2003, stolen from the Iraq Museum or looted from illegal digs. Now some, at least, are on the road back to Iraq. Syrian officials in April seized 700 artifacts from smugglers and dealers and sent them back to Iraq. Just last week, Jordan announced the repatriation of 2400 artifacts seized by customs authorities in antismuggling operations. John Russell, an archaeologist at the Massachusetts College of Art in Boston who is consulting with the U.S. State Department, says that about 1000 artifacts—including tablets, cylinder seals, and glass bottles—intercepted by customs officials will be turned over soon to the Iraq Embassy in Washington, D.C.

Meanwhile, efforts are under way to restore Iraq's fraying network of regional museums. Italy is working to rehabilitate several, and Curtis says the British military may provide $20 million to convert one of Saddam's palaces in Basra into a museum.

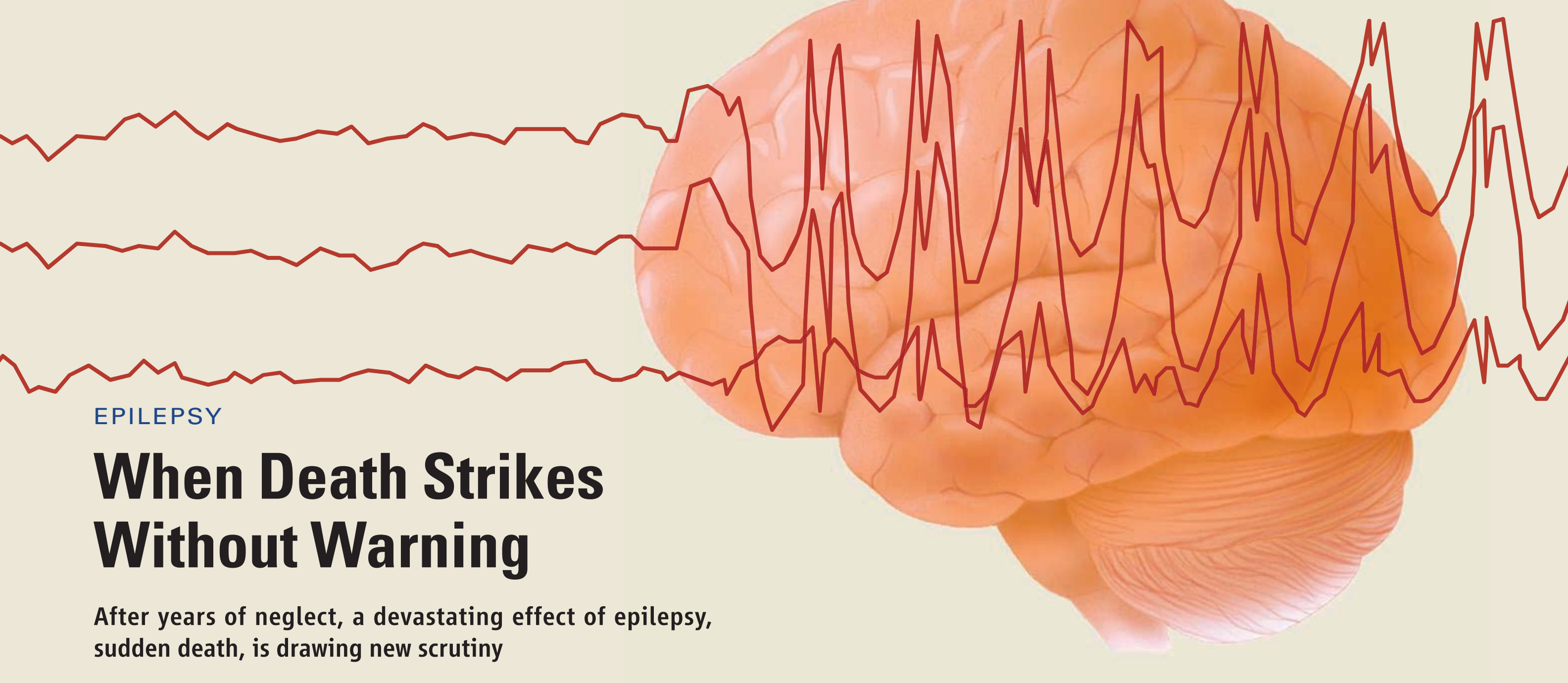
One of the most frustrating tasks confronting archaeologists concerns the ancient capital of Babylon, located 85 kilometers south of Baghdad. It was a major player in Middle Eastern history from the 23rd century B.C.E. until just before the time of Christ. American and Polish troops damaged parts of the metropolis while building a military base there, according to a 2005 report by Curtis. As an exhibit on the city's historical impact opened in Berlin last week, researchers from around the world gathered nearby to discuss how to manage the decaying site and stave off plans for development, including new parking lots and a hotel. The rescue effort has been stymied for years by changes in the archaeology leadership in Baghdad and bickering between Americans and Europeans. "It is very disappointing it has taken so long to agree on an assessment," says Curtis, who visited the site in 2004. "Only after that is done can we move forward." The U.S. government intends shortly to announce a $700,000 contract with the World Monuments Fund to begin work on the management plan.

Archaeologists may learn more during an upcoming U.S.-sponsored visit to sites in central or northern Iraq, according to spokesperson Olson and archaeologist Diane Siebrandt, also of the State Department in Baghdad. Olson declined to discuss the trip, however, citing "operational considerations," and Siebrandt would not provide details about any U.S. efforts to cope with the damage and looting resulting from the war.

Stone, meanwhile, sees a silver lining in the havoc. The focus on satellite data may help archaeologists unable to work on the ground understand ancient Mesopotamian settlement patterns and architecture, she says, gaining fresh insight into how its inhabitants once lived. **–ANDREW LAWLER**

CREDIT: ABDEL-AMIR HAMDANI/SBA

Published by AAAS



EPILEPSY

When Death Strikes Without Warning

After years of neglect, a devastating effect of epilepsy, sudden death, is drawing new scrutiny

The call came on a Thursday, 21 February 2002, while Jeanne Donalty sat at her desk at work. Her son Chris, a 21-year-old senior at a Florida college, had stopped breathing. His girlfriend found him on his bed, surrounded by the books he'd been studying and a summer job application. Paramedics were unable to revive him, and just like that, Chris Donalty was gone.

Chris Donalty had had epilepsy—he suffered his first seizure in school when he was 9 years old—but his mother at first saw no clear line connecting his death and the disease for which he was being treated. An autopsy found no visible cause of death, and it was shortly after that that Jeanne Donalty discovered a term she had never heard before: SUDEP.

Sudden unexpected death in epilepsy, SUDEP was first written up in *The Lancet* in 1868 by a British physician; he described the phenomenon as "sudden death in a fit." Neurologists today are familiar with SUDEP, which is thought to follow a seizure, and most specialists have lost patients in this way. "Four or five times a year, someone will not come to my clinic because they have a SUDEP death," says Mark Richardson, a neurologist at King's College London. Most victims, like Chris Donalty, are in their 20s or 30s.

SUDEP has been little studied and is rarely discussed in the medical and scientific communities. Families often learn of it only after a relative's death. In the United Kingdom, which is well ahead of the United States in tracking SUDEP, it's estimated that SUDEP strikes at least 500 people a year. It's thought to explain between 8% and 17% of deaths in people with epilepsy. Among those with frequent seizures, the number may be as high as 40%. This increased risk, recognized only recently, underscores that SUDEP is more likely to occur if seizures are more frequent or treatment is inadequate.

Chris Donalty was in that high-risk group: Despite taking his medications as prescribed, he suffered seizures regularly for 2 years before his death. But he never told his parents—because, they now believe, he did not want to lose his driver's license. "I don't know of any other disease that can be fatal where patients aren't aware of" that risk, says his mother.

Driven largely by grieving families, more doctors are discussing risk of SUDEP with patients, and research is picking up. A few studies are focusing on what happens to breathing and heart rhythm during seizures. In the U.K., researchers and advocates hope to set up a nationwide registry of SUDEP cases. The U.S. National Institutes of Health (NIH) will host several dozen specialists at its Bethesda, Maryland, campus this fall in a first-ever meeting on SUDEP. Still, the epilepsy community is divided on what to tell patients about the risk of sudden death—and exactly what should be done about it.

Detective. Neurologist Lina Nashef dug into historical data in a quest to understand what causes sudden death in epilepsy.

In from the shadows

Epilepsy, characterized by recurrent seizures caused by abnormal electrical activity in the brain, has long carried a stigma. Some say this may explain why physicians swept SUDEP under the rug: They didn't want to magnify existing fears, especially because no way to prevent it is known. "There was a real concern that the main message should be, 'You can live a completely normal life with epilepsy,' " says Jane Hanna, who helped found the nonprofit Epilepsy Bereaved in Wantage, U.K., after her 27-year-old partner died of SUDEP shortly after he was diagnosed. Even textbooks on epilepsy omitted mention of SUDEP.

But this discretion carried drawbacks, burying historical knowledge of SUDEP cases and slowing clinical investigation, says Lina Nashef, a neurologist at King's College Hospital in London. Until the early 20th century, many people with epilepsy lived in asylums or other institutions, where staff recognized that patients sometimes died during or after seizures. But the collective memory of these deaths faded as antiepilepsy drugs became widely available and patients began living independently. Most who die of SUDEP now do so at home, unobserved.

Nashef began investigating SUDEP as a research project for a postgraduate degree in 1993,

CREDITS (TOP TO BOTTOM): K. SUTLIFFE/SCIENCE; RINA PATEL/KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST

interviewing 26 families who had lost someone to sudden death. Although nearly all the deaths occurred without witnesses, Nashef was often told of signs, such as a bitten tongue, that occur after a seizure. The evidence in other cases was more circumstantial: One young man in his late teens, whose seizures were triggered by flickering light from television and computer screens, was found dead at a computer terminal in the library.

Nashef identified a handful of characteristics that the SUDEP victims shared. All but three were battling regular seizures, though sometimes not more than two or three a year. And all had suffered from a particular type, called generalized tonic-clonic or, colloquially, grand mal seizure. Such seizures, the kind most people associate with epilepsy, are accompanied by a loss of consciousness and violent jerking motions and affect large swaths of the brain.

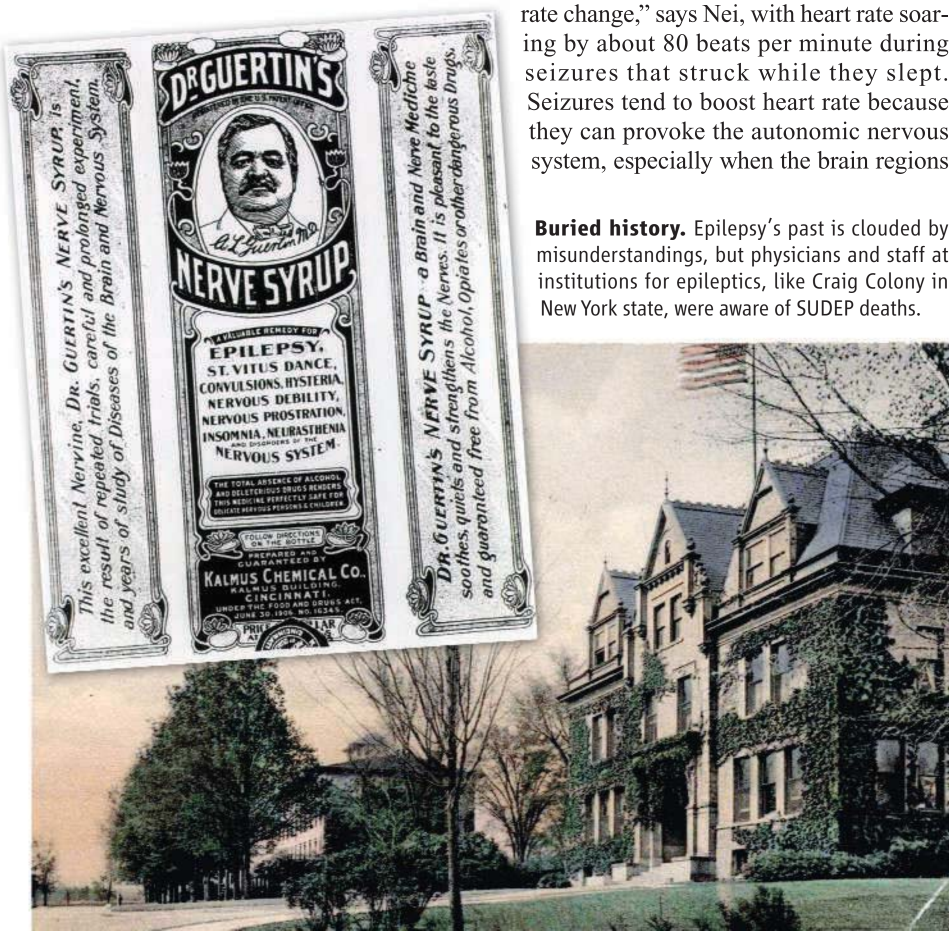
What goes wrong?

Digging deeper into SUDEP, Nashef and others have focused on two life-sustaining functions: respiration and heartbeat. Most physicians now believe that SUDEP stems from arrested breathing, called apnea, or heartbeat, called asystole.

One broader question is whether apnea or asystole strike even during seizures that aren't fatal. Neurologists Maromi Nei and her mentor, Michael Sperling, both at Thomas Jefferson University in Philadelphia, provided an early clue in 2000 when they described electrocardiogram patterns from 43 people with epilepsy. Although none died of SUDEP, 17 of these patients had cardiac abnormalities during or right after seizures, including significant arrhythmias and, in one case, no heartbeat at all for 6 seconds.

> ***"Four or five times a year, someone will not come to my clinic because they have a SUDEP death."***
>
> **—MARK RICHARDSON, KING'S COLLEGE LONDON**

More recently, Nei and her colleagues investigated hospital records from 21 people who later died, apparently of SUDEP, and compared their heart rhythms with those from the original study, to see whether the SUDEP cohort had some signs of susceptibility. The biggest difference, they reported in 2004, was not the prevalence of arrhythmias but "a greater degree of heart rate change," says Nei, with heart rate soaring by about 80 beats per minute during seizures that struck while they slept. Seizures tend to boost heart rate because they can provoke the autonomic nervous system, especially when the brain regions stimulated are those that trigger such "fight-or-flight" reactions. These data hinted that the phenomenon is exaggerated in those who later die of SUDEP.


DR. GUERTIN'S
REGISTERED IN THE U.S. PATENT OFFICE
A. L. Guertin M.D.
NERVE SYRUP
A VALUABLE REMEDY FOR
EPILEPSY,
ST. VITUS DANCE,
CONVULSIONS, HYSTERIA,
NERVOUS DEBILITY,
NERVOUS PROSTRATION,
INSOMNIA, NEURASTHENIA
AND DISORDERS OF THE
NERVOUS SYSTEM.
THE TOTAL ABSENCE OF ALCOHOL AND DELETERIOUS DRUGS RENDERS THIS MEDICINE PERFECTLY SAFE FOR DELICATE NERVOUS PERSONS & CHILDREN
FOLLOW DIRECTIONS ON THE BOTTLE
PREPARED AND GUARANTEED BY
KALMUS CHEMICAL CO.
KALMUS BUILDING,
CINCINNATI.
UNDER THE FOOD AND DRUGS ACT, JUNE 30, 1906. NO. 16345.
This excellent Nervine, DR. GUERTIN'S NERVE SYRUP, is the result of repeated trials, careful and prolonged experiment, and years of study of Diseases of the Brain and Nervous System.
DR. GUERTIN'S NERVE SYRUP a Brain and Nerve Medicine soothes, quiets and strengthens the Nerves. It is pleasant to the taste and guaranteed free from Alcohol, Opiates or other dangerous Drugs.

Buried history. Epilepsy's past is clouded by misunderstandings, but physicians and staff at institutions for epileptics, like Craig Colony in New York state, were aware of SUDEP deaths.

Now Nei is implanting devices under the left collarbone of 19 people with intractable epilepsy to gather data on their heart rhythm over a span of 14 months. Neurologist Paul Cooper of Hope Hospital near Manchester, U.K., is beginning a similar study with 200 people.

Both studies follow a related and troubling report in 2004 from *The Lancet*. There, a group of British researchers described cardiac data from 377 seizures in 20 patients gathered over 2 years. Four of the 20 had perilous stretches of asystole and later had pacemakers permanently implanted to jump-start their hearts if needed.

What might be behind this effect? Asystole isn't always dangerous, although it sounds alarming; it can happen even during some fainting spells. A normal heart starts beating again on its own—which leads clinicians to wonder whether the hearts of patients struck by SUDEP may harbor invisible defects. One possibility is that over time, repeated seizures can scar and damage the organ. Another is that a genetic defect may be causing both heart rhythm problems and epilepsy.

Earlier this year, Nashef, King's College geneticist Neeti Hindocha, and their colleagues intrigued epilepsy specialists with a report on a family with a rare form of inherited epilepsy, including two members who died from SUDEP. The researchers, after gathering DNA from the living, found that all 10 family members who had epilepsy also carried a previously undescribed mutation in a gene called *SCN1A*, which was responsible for their disease. A so-called ion channel gene, *SCN1A* helps control electrical signaling between cells. Similar genes have been linked to epilepsy and sudden cardiac death. The authors postulated that the SUDEP deaths in this family were also caused by *SCN1A*, which could have disrupted heart rhythm or brainstem function in addition to triggering epilepsy. A group at Baylor College of Medicine in Houston, Texas, is now studying whether ion channel genes that can freeze the heart are also present in brain tissue.

If cardiac defects like these are behind SUDEP, "it might be something preventable," says Stephan Schuele, director of the Com-

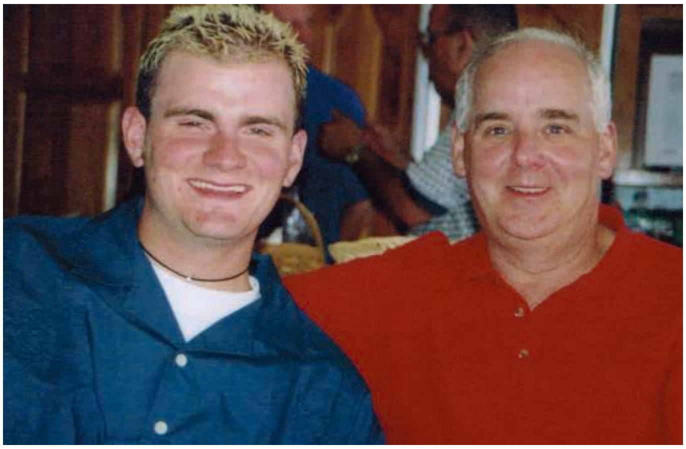
CREDITS (TOP TO BOTTOM): NATIONAL LIBRARY OF MEDICINE; COURTESY OF THE MUSEUM OF DISABILITY HISTORY

Published by AAAS

prehensive Epilepsy Center at Northwestern Memorial Hospital in Chicago, Illinois. People with gene defects that cause sudden cardiac death, for example, receive pacemakers that can shock their hearts into beating again. Perhaps, doctors say, the same could be done for epilepsy—if they can determine who's at risk of SUDEP to begin with.

Missing clues

But Schuele, who's looking for other causes of SUDEP, notes that despite a few reports pointing to genetics, "there is no direct evidence" that asystole is killing people with epilepsy. Schuele wonders if the body's way of stopping seizures in the brain could also be disturbing vital brainstem function in some patients. These mechanisms, which are just starting to be explored and involve surges of certain neurotransmitters, may go overboard and cause chaos in the autonomic nervous system, which governs heart rate and respiration.



SUDEP victim. Chris Donalty, shown here with his father, Barry, died after a seizure at age 21 in his senior year of college.

CREDIT: TRISHA DONALTY

The detective work is slow and arduous, in part because so few cases of SUDEP have come to light from epilepsy monitoring units in hospitals, where vital signs are recorded—perhaps, Schuele suggests, because health workers are loath to admit that a SUDEP death occurred on their watch. Last August, neurologists Philippe Ryvlin of the Hospices Civils de Lyon in France and Torbjörn Tomson of the Karolinska Hospital in Stockholm, Sweden, began surveying 180 hospitals in Europe for information on SUDEP deaths or "near-misses" that required resuscitation. Two months ago, they extended their search worldwide, collecting cases from as far away as India and the United States. They expect to conclude their collection and analysis in about a year.

Just four cases have been published. The most detailed, in 1997 on a patient in Bristol, U.K., reported that that person's brain waves went flat before the pulse faded, perhaps causing a failure of the brain region that controls breathing. This suggests that heart failure could be a consequence, not a cause, of SUDEP. Still, "the mechanism of that brain-activated shutdown is very mysterious," says Ryvlin. "Nobody knows what it could be."

There are clues that respiration is key. By monitoring it in hospitalized epilepsy patients, Nashef found that episodes of apnea were common during seizures. And a mouse strain used for decades to test epilepsy drugs has the disconcerting habit of dying from respiratory failure after a severe seizure. That was "generally considered a nuisance," says Carl Faingold, a neuropharmacologist at Southern Illinois University in Springfield, until he and a handful of others realized the mice could be used to study SUDEP. At Boston College, biologist Thomas Seyfried found that putting the mice in an oxygen chamber during seizures prevented death in all of them.

Faingold considered whether the neurotransmitter serotonin, which functions in the brain's respiratory network, might play a role. He gave the mice the antidepressant Prozac, a serotonin booster, and found that though their seizures remained the same, they were at least 90% less likely to die afterward.

Faingold is disappointed that the mouse work has received little attention and no financial support from NIH—his SUDEP research is funded by an epilepsy advocacy group—and its relevance to humans has been questioned. Because no one can predict who will die of SUDEP or when, "if you don't have a way of investigating [SUDEP] in animals, you're very limited," he says.

Acknowledging the unmentionable

Meanwhile, doctors face a more pressing question: what to tell their patients about SUDEP. "I admit, I am still trying to figure out the best way to do this," says Elizabeth Donner, a pediatric neurologist at the Hospital for Sick Children in Toronto, Canada. She has grown more willing to share the information, but still, "sometimes we worry in telling people about this phenomenon … we could actually make their lives worse." Already, one of the toughest aspects of epilepsy is its unpredictability. "When you add in a statement that some people die, and we don't know why and we can't predict it and we can't prevent it, that can be very scary."

U.K. national guidelines in 2004 recommended that physicians discuss SUDEP with everyone who has epilepsy. In reality, a survey of British neurologists published 2 years ago showed, "nobody told anybody anything," says Cooper. Cooper and some other physicians believe that the 30% or so of patients whose epilepsy does not respond to medication—or those reluctant to take it—ought to be told of SUDEP, because they are at a higher risk than people whose epilepsy is controlled. The latter group, he believes, does not need to know about SUDEP.

That perspective doesn't sit well with epilepsy advocates. "Anecdotally, we're aware of deaths every year in people with second or third seizures," says Hanna of Epilepsy Bereaved. "It does worry me a bit if there's going to be some basic clinical practice that just cuts the line with people who seem to have the most serious epilepsy."

Jeanne Donalty still struggles with her family's ignorance of SUDEP while Chris was alive. "I'm not insensitive to how hard this is for a physician," she says. If she had known of SUDEP then, "I would have been upset; … who wouldn't be? But I think you have the right to have all the knowledge about the disease that is out there, so that you can make your decisions based on that knowledge." When it comes to sharing information on SUDEP, says Donalty, "to me, it's easy. You tell everybody."

–JENNIFER COUZIN

Nearly departed? Parts of SLAC's PEP-II collider could be shipped to Italy to build a new collider for high-precision experiments, called SuperB.

CREDIT: SLAC

PARTICLE PHYSICS

Competing Teams Plot Two Different Paths to a New Particle Smasher

To make a new collider, physicists in Japan plan to push an existing machine to its limits. Others in Italy hope to cobble one together from old parts and a bright idea

Many a teenager has dreamed of transforming a jalopy into a gleaming hot rod. Now, a team of physicists from the United States and Italy has proposed a project that sounds as unlikely. Using parts from an old particle smasher, they plan to build a new one that will crank out data 100 times faster than the original machine, consume less power, and possibly find hints of particles so massive that no collider could produce them directly—not even the new highest energy collider that will turn on in Europe this summer. But the project, dubbed SuperB, isn't the only dragster in this race: Physicists in Japan plan to upgrade their existing machine to do the same work.

SuperB would be built at the University of Rome "Tor Vergata," near Frascati National Laboratory in central Italy. But most of its parts would come from the PEP-II collider at the Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, which was shut down in April—even though some say it still had plenty of science in it. SuperB team members hope SLAC and the U.S. Department of Energy (DOE) will donate PEP-II and the accompanying BaBar particle detector to the project as an in-kind contribution worth about $200 million. "Here's a contribution that doesn't cost anybody anything," says David Hitlin, a team member from the California Institute of Technology (Caltech) in Pasadena. "Doesn't it make sense to leverage your assets?"

SuperB would serve as a foil to the world's mightiest accelerator, the Large Hadron Collider (LHC) soon to power up at the European particle physics laboratory, CERN, near Geneva, Switzerland (*Science*, 23 March 2007, p. 1652). By smashing protons into protons, the LHC aims to blast massive new particles into existence. In contrast, SuperB would collide electrons and positrons at lower energies to produce a flood of familiar particles, and the details of their decays could reveal hints of new physics.

The approach, called precision physics, has the potential to be "real cowboy physics," says Thomas Browder of the University of Hawaii, Honolulu. Such a collider might spot rare decays that would rewrite the standard model of particle physics or even find hints of particles beyond the grasp of the LHC, Browder says.

But SuperB has competition. Browder is one of about 400 physicists working with the KEKB collider and the Belle particle detector at the Japanese laboratory KEK in Tsukuba. They plan to upgrade that machine to create Super KEKB. "This was put into the official plan of KEK" in January, says Masanori Yamauchi, a particle physicist at KEK, "but the government has not given approval yet."

Heavy hints

Because mass and energy are equivalent, physicists can pop a massive new particle into existence by colliding well-known ones at sufficiently high energy, as they aim to do at the LHC. But massive new particles can also cast shadows in the decays of far less massive ones, especially those made up of fundamental bits of matter called quarks.

According to the standard model of particles, the matter around us consists of the up quarks and down quarks that make up protons and neutrons, electrons, and wispy electron neutrinos. This first "family" of particles is copied twice over, so there are heavier quarks of four more "flavors": charm and strange, top and bottom.

Consider the decay of a particle called a B meson, which contains a massive bottom quark and a lighter antiquark. Thanks to the uncertainties of quantum mechanics, the meson roils with other particles popping in and out of "virtual" existence within it, even ones more massive than the meson itself. So if there are new particles lurking over the horizon, they will flit about inside the meson and may reveal their nature by affecting the way the B meson decays.

Physicists have used this approach to narrow in on new particles before. For example, in the 1980s, studies of B mesons, which are only five times as heavy as a proton, indicated that the then-hypothesized top quark was much heavier than previously thought, says Peter Krizan of the University of Ljubljana and the Jožef Stefan Institute in Slovenia. That inference proved correct when the top quark was found in 1995 and weighed in at 180 times the mass of a proton.

Both the KEKB collider and SLAC's PEP-II were built to do just this sort of work. Since 1999, the two "B factories" have pumped out scads of B mesons, and experimenters working with the BaBar detector at SLAC and the Belle detector at KEK have studied a slight asymmetry between B mesons and their antimatter counterparts, anti–B mesons. That discrepancy, known as charge-parity (CP) violation, had been previously seen only in lighter K mesons.

BaBar and Belle proved that, to a precision of a few percent, the standard model's explanation of CP violation is on the mark (*Science*,

Published by AAAS

13 October 2006, p. 248). That was both a huge victory and a disappointment for physicists, as the theory contains far too little CP violation to explain why the universe contains gobs of matter but essentially no antimatter. "We all know that the standard model is a fantastic theory," Krizan says, "but we also know that it's fantastically wrong."

Both the SuperB and KEKB teams now want a "super flavor factory" that will crank out far more B mesons, as well as mounds of particles called D mesons and tau leptons, heavier cousins of electrons. All that data would allow for even more precise tests of the standard model's CP-violation scheme. More important, says Hitlin, it might reveal rare decays that turn the theory on its head. "The point is *not* doing what you did before but better," Hitlin says. "It's looking for these very rare decays."

Such studies would complement the LHC's direct search for new particles. If the LHC sees plenty of new particles, a super flavor factory would probe how they couple to quarks and other known particles. If the LHC sees nothing, then precision physics offers the best hope of sensing particles beyond its grasp. "These precision measurements are basically the only tool you have that shoots far beyond the mass reach of the LHC," Krizan says.

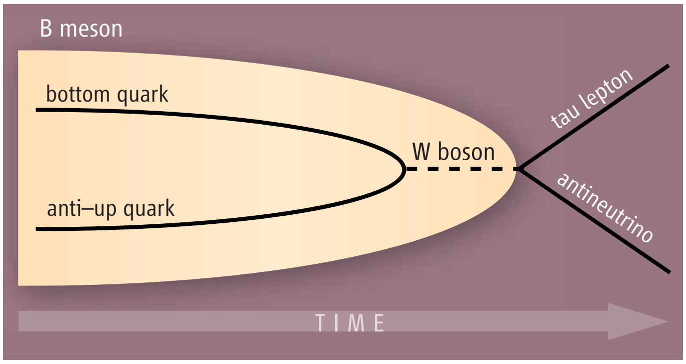
Huge currents, tiny beams

The SuperB and KEK groups are taking different approaches to designing their machines. Similar to PEP-II, the KEKB collider comprises two circular accelerators that cross in the middle of the associated detector, one carrying electrons in one direction and the other carrying positrons in the other. "Our design is kind of brute force," says Yamauchi. "We put more and more electrons and positrons into the rings."

KEK physicists would boost the current in the electron ring from 1.2 amps to 4.1 amps and in the lower energy positron ring from 1.6 amps to a sizzling 9.4 amps. They would squeeze the beams to half their current size and employ a new technique to reduce the tendency of the crossing beams to disrupt each other. The path to Super KEKB is "very, very predictable from our present machine," says Katsunobu Oide, an accelerator physicist at KEK. By the time KEKB shuts down, probably in 2010, it will have created a billion B–anti-B pairs. Super KEKB would produce pairs at least 10 times faster and eventually make 50 billion of them.

Instead of packing in more particles, SuperB would use greatly compressed beams, thereby increasing the rate at which electrons and positrons collide, which is called the luminosity. "We get 100 times smaller vertical size at the interaction point, and that means 100 times more luminosity with the same beam current," says Pantaleo Raimondi, an accelerator physicist at Frascati who dreamed up the scheme. At the start, SuperB would crank out data five times as fast as Super KEKB's initial rate.

SuperB would collide beams only 35 nanometers across. To make such tiny beams, researchers must very precisely arrange both the magnets that steer the beam around a ring and those that focus it, Raimondi says. The SuperB design borrows from work on "damping rings" being developed to compress the beams in the proposed International Linear Collider (ILC), a multibillion-dollar straight-shot collider that would study in detail new particles discovered at the LHC.



Subtle signals. A B meson decays into a tau lepton and an antineutrino. The probability for the decay would differ from standard model predictions if there are new particles that could fill the role of the familiar W boson.

To limit the cost of SuperB to roughly $500 million, researchers plan to reuse the PEP-II hardware from SLAC. In fact, physicists had proposed upgrading PEP-II where it stands as early as 2001. Those plans were squeezed out by tight budgets in DOE's particle physics program and by the U.S. community's desire to push to host ILC. Now that a PEP-II upgrade is "not in the cards," the lab may be willing to part with the machine, says Steven Kahn, SLAC's director of particle physic and astrophysics. "We're not seeing any major hurdles to our saying yes to this," he says. SLAC has asked the Italian National Institute for Nuclear Physics (INFN) to formally request the equipment, Kahn says.

Pros and cons

Each approach has both strengths and potential weaknesses. The Super KEKB design requires no conceptual leaps, but circulating nearly 10 amps of current presents its own challenges. The extent to which the beams disrupt each other increases with the number of particles in them, says John Seeman, an accelerator physicist at SLAC, so achieving the luminosity increase may be tricky. The high currents would also increase power consumption of the complex from 40 megawatts to 80 megawatts, raising yearly operating costs by tens of millions of dollars.

In contrast, the SuperB collider would use only 20 megawatts, less than PEP-II did. But steering its tiny beams into each other may be tough, Oide says. "To collide such tiny beams is not trivial," he says. "It's many orders of magnitude more difficult than producing a single nanometer-sized beam." SuperB researchers will have to limit vibrations at the crossing point to just 3 nanometers, Oide says. However, if the tiny-beam scheme seems likely to work, then KEK researchers may simply adopt it, too.

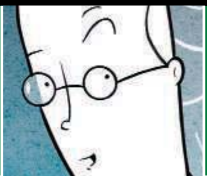
Politically, SuperB team members have a tougher row to hoe, as they are asking the Italian government for hundreds of millions of euros to build a new laboratory to house the collider. A subpanel of the European Committee for Future Accelerators is studying the plan. If both it and the CERN Strategy Group, which keeps the road map for European particle physics, give the plan high marks, then INFN will ask the Italian government for funding. Physicists hope to begin detailed design work as early as next year.

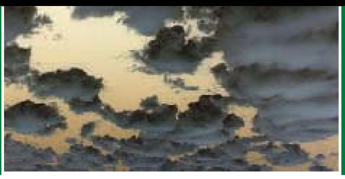
In contrast, KEK researchers already have a lab and machine. KEK is negotiating for funding with Japan's Ministry of Education, Culture, Sports, Science, and Technology. Researchers hope to shut down KEKB in 2010 and spend 3 years building Super KEKB. At the least, they hope to use the money saved from KEKB's operating budget to fund $220 million in improvements. The full upgrade would cost much more, but Japanese researchers are reluctant to say how much.

Given the financial demands of the LHC and other projects and tight funding all over, many say the community can likely afford only one super flavor factory. "In the end, the country that wants the machine the most and puts up most of the money will get it," Seeman predicts. Will it be Italy or Japan? Physicists may know within a year.

–ADRIAN CHO

Published by AAAS





LETTERS

edited by Jennifer Sills

Painful Publishing

BIOMEDICAL SCIENCE HAS NEVER BEEN MORE EXCITING OR PRODUCTIVE. RESEARCH TOOLS have become increasingly powerful, and progress continues to accelerate. Yet, these are stressful times for many biomedical scientists, because competition for grant support, jobs, and publishing in the most prestigious journals is also accelerating. The stress associated with publishing experimental results—a process that can take as long as obtaining the results in the first place—can drain much of the joy from practicing science.

One problem with the current publication process arises from the overwhelming importance given to papers published in high-impact journals such as *Science*. Sadly, career advancement can depend more on where you publish than what you publish. Consequently, authors are so keen to publish in these select journals that they are willing to carry out extra, time-consuming experiments suggested by referees, even when the results could strengthen the conclusions only marginally. All too often, young scientists spend many months doing such "referees' experiments." Their time and effort would frequently be better spent trying to move their project forward rather than sideways. There is also an inherent danger in doing experiments to obtain results that a referee demands to see. Although we emphasize these problems with regard to the highest-impact journals, the same problems occur with other journals.

"The stress associated with publishing experimental results…can drain much of the joy from practicing science."

It is surprising that so many referees make unnecessary demands, as they are authors themselves and know how it feels when the situation is reversed. Such demands are discouraging for young scientists and, cumulatively, slow the progress of science. Of course, peer review is critical for making sure that the authors' conclusions are sound, and some referees' experiments would substantially advance the story. But frequently, these would justify an additional paper. Science advances in stages, and no story is complete.

What can be done to speed up the publication process and make it less agonizing and more efficient? Both editors and referees could help. Referees need to be more thoughtful when recommending additional experiments and to make sure that these experiments are truly needed to justify publication. Editors should insist that reviewers rigorously justify each new experiment that they request. They should also ask reviewers to estimate how much time and effort the experiment might require. With this information in hand, editors can more easily override referees' excessive demands. This requires confident, knowledgeable, and experienced editors, and it risks alienating referees, who are often hard to come by. Nonetheless, editors should be encouraged and empowered to perform this crucial task.

A more radical solution, which is already used by some journals, is to have editors and their relevant editorial board members triage papers so that only those that meet the criteria of interest, novelty, and importance appropriate for the journal are sent out for formal review. This will save reviewers' time. In addition, papers that clear this initial hurdle can then be reviewed solely for scientific accuracy, appropriateness of controls, clear writing, and justification of the conclusions.

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Published papers are the currency of science, and scientists need to do more to make the publishing process more rapid, rational, and equitable, as well as less painful and frustrating. We scientists have created the problems discussed here, and it is up to us to fix them.

MARTIN RAFF,[1] ALEXANDER JOHNSON,[2] PETER WALTER[3]

[1]Emeritus Professor, Department of Biology, University College London, London WC1E 6BT, UK. [2]Department of Microbiology and Immunology, University of California, San Francisco, CA 94158, USA. [3]Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA.

The Enemy Within

THE NEWS OF THE WEEK STORY BY D. GRIMM on "Staggering toward a global strategy on alcohol abuse" (16 May, p. 862) nicely illustrates the uphill battle that the World Health Organization faces in dealing with global health issues. I was dismayed (but not surprised) to learn that several countries (including the United States) insisted that the Director General of WHO include the alcohol industry in discussions to shape global strategy concerning alcohol abuse. The alcohol industry is equivalent to the "vector" for alcohol-induced disease. Inviting this industry to the discussion table regarding attempts to curb alcohol-related deaths is analogous to inviting the mosquito to participate in discussion concerning the control of malaria.

VICTOR L. ROGGLI

Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA. E-mail: roggl002@mc.duke.edu

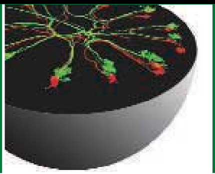
The Limits of Water Pumps

WATER STRESS IS A MAJOR PROBLEM AFFECTing the future of human societies around the world, particularly in the rural areas of the developing world [1, 2]. The Newsmakers article "Barren to lush" (2 May, p. 593) highlighted an award for the invention of a new manual pump used for irrigation in rural Africa. We fully respect and admire the invention of efficient and affordable pumping systems to solve water-shortage problems in rural areas. However, we are concerned about the intensive application of these new pumps to water-limited systems, where the extraction of

Published by AAAS

Deciding under influence

47



Transience in neuronal networks

48

groundwater and its use in agriculture could be unsustainable, despite the recognition of this new technology with an Award for Sustainability (as noted in the Newsmakers item). The irrigation pump will undoubtedly bring short-term benefits, but it could have adverse long-term consequences. First, groundwater pumping can deplete the limited groundwater. Second, the extraction of groundwater and its use for irrigation increases soil evaporation, which, in turn, may increase soil salinity and unproductive water losses. Third, in coastal areas groundwater pumping causes seawater intrusion. All of these situations are examples of how intensive groundwater extraction in areas with only limited recharge rates may lead to an unsustainable use of the landscape.

This is also true in many pastoralist societies, where the increase in water availability often leads to the overgrazing of rangelands. The case of Botswana is representative of other rural parts of Africa. For example, in the Kgalagadu District, the number of boreholes increased from 8 in the 1950s to more than 380 in the 1990s (*3*), resulting in higher rates of livestock production, overgrazing, and consequent land degradation.

We think pumps are good for solving short-term drinking water shortages. However, new technology aiming at solving the long-term agricultural water shortage in rural regions should focus on more efficient use of natural rainfall (e.g., efficient rainfall collectors and reduction of soil evaporation) or wastewater reuse. In this way, science and new technologies can move in the same direction.

LIXIN WANG[1,2] AND PAOLO D'ODORICO[2]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA. [2]Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA.

References

1. J. Yeston, R. Coontz, J. Smith, C. Ash, *Science* **313**, 1067 (2006).
2. A. Tal, *Science* **313**, 1081 (2006).
3. D. S. G. Thomas, C. Twyman, *Land Degradation Dev.* **15**, 215 (2004).

Omissions in GLAST Story

IN THE NEWS FOCUS STORY "GLAST MISSION prepares to explore the extremes of cosmic violence" (23 May, p. 1008), Y. Bhattacharjee committed two grave oversights.

First, no mention is made of any contribution to GLAST from outside of the United States. In fact, Italy, France, Sweden, and Japan all made essential contributions. The Large Area Telescope, for example, was essentially made and paid for by Italy, France, and Sweden. Japan supplied most of the necessary silicon. Scientists from these countries have been, and continue to be, essential members of the GLAST team.

Second, the figure on page 1009, which provides a brief summary of high-energy astronomy missions, omits two important missions: Italy's (and Holland's) BeppoSAX (1996 to 2002) and Italy's AGILE. BeppoSAX has substantially added to our understanding of gamma-ray bursts and hard x-rays; the BeppoSAX team was awarded the 1998 Bruno Rossi Prize of the American Astronomical Society. AGILE is also dedicated to x-ray and gamma-ray astronomy and uses the same silicon type of detectors that GLAST will use. Now in orbit for more than a year, AGILE is certainly a precursor to (and pathfinder for) GLAST.

As a final clarification, in the same figure, the Swift mission is a joint trilateral mission with NASA, Italy, and the UK, not NASA alone, as indicated.

GIOVANNI F. BIGNAMI,[1] T. MACCACARO,[2] R. PETRONZIO,[3] M. TAVANI[4]

[1]Italian Space Agency (ASI), 00198 Rome, Italy. [2]Italian Institute for Astrophysics, 00136 Rome, Italy. [3]Italian Institute for Nuclear Physics, 00044 Rome, Italy. [4]AGILE, Italian Institute for Astrophysics, 00136 Rome, Italy.

LIFE IN SCIENCE

Frogs on a Plane

In the early days, one of the subjects of our research was the *Engystomops* frog, a tiny creature known for its brown pustular skin (*1*). When the lab moved from New York City to California in 1971, a young assistant named Bill was entrusted with transporting the frog colony to its new home. After painstakingly sifting through all the dirt in our "Little Panama" culture room, he placed the frogs in aquaria. He decided that it would be safer to carry the frogs onto the airplane with him than to trust them to checked luggage. So the morning of the flight, he carefully put the frogs into plastic bags with water and air, and then placed each bag into his carry-on suitcase. Unfortunately, despite his meticulous planning, there was one thing he forgot to take into account.

As soon as the plane took off, the change in air pressure caused the bags to burst. Of course, Bill couldn't help opening the carry-on to see how bad the situation was. When he saw what had happened, he asked every flight attendant he could find for glasses of water that he could use to refill and retie the bags. But he was too late: Out jumped the frogs. Bill and the startled flight attendants raced around the plane, crawling under seats and down the aisles to apprehend the little creatures. Baffled passengers looked on, trying to determine the source of the commotion.

Fortunately, there was a happy ending to this little adventure. Eventually, the frogs were caught and transported safely to our California lab, where they would prosper for many years to come.

EDITOR'S NOTE

This will be an occasional feature highlighting some of the day-to-day humorous realities that face our readers. Can you top this? Submit your best stories at www.submit2science.org.

JANE RIGG

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

Reference

1. Search for *Engystomops pustulosus* on http://amphibiaweb.org; don't miss the mating call!

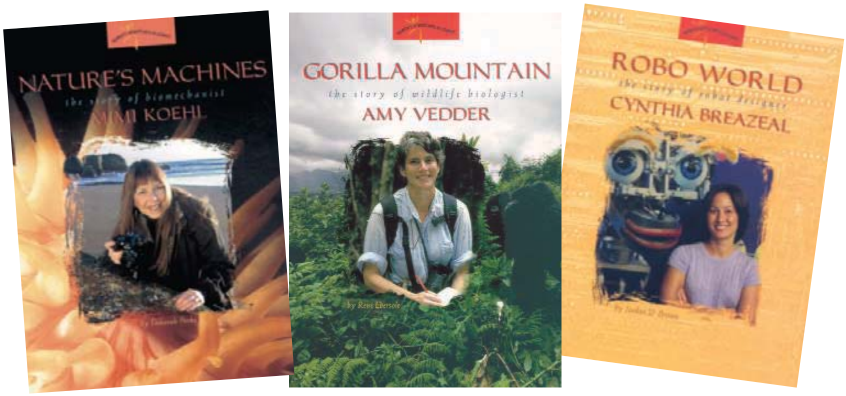
CREDIT: PETER HOEY

Suggestions for Your Summer Reading Enjoyment

YOUTH NONFICTION

Inspiring Experiences in Science

Susan Kovats


NATURE'S MACHINES
the story of biomechanist
MIMI KOEHL
GORILLA MOUNTAIN
the story of wildlife biologist
AMY VEDDER
ROBO WORLD
the story of robot designer
CYNTHIA BREAZEAL

Gorilla Mountain. The Story of Wildlife Biologist Amy Vedder. *Rene Ebersole.* Franklin Watts (Scholastic), New York, and Joseph Henry (National Academies Press), Washington, DC, 2005. 128 pp. $9.95. ISBN 9780309095518.

Robo World. The Story of Robot Designer Cynthia Breazeal. *Jordan D. Brown.* Franklin Watts (Scholastic), New York, and Joseph Henry (National Academies Press), Washington, DC, 2005. 128 pp. $9.95. ISBN 9780309095563.

Nature's Machines. The Story of Biomechanist Mimi Koehl. *Deborah Parks.* Franklin Watts (Scholastic), New York, and Joseph Henry (National Academies Press), Washington, DC, 2005. 128 pp. $9.95. ISBN 9780309095594.

Mimi Koehl's "I had to know the answer!" sums up the dedication, effort, and excitement that the women scientists featured in *Gorilla Mountain*, *Robo World*, and *Nature's Machines* (respectively, Amy Vedder, Cynthia Breazeal, and Mimi Koehl) bring to their research projects. The three titles are part of the outstanding ten-book series Women's Adventures in Sciences, underwritten by the U.S. National Academy of Sciences, which features the lives and research areas of contemporary women scientists (*1*).

Engaging, inspiring, and informative, the books describe the childhood, scientific training, research topics and careers, and personal family lives of the scientists. Although written for middle school and secondary students, the books may be read at multiple levels. Thus they should appeal to the intended audience, and they are also likely to capture the attention of younger children and adults. Photographs, which appear on nearly every page, depict research materials and equipment; laboratories or field locations; scientists working; and families, mentors, or colleagues. These pictures both personalize the stories and show how the research is actually accomplished. Boxed inserts such as "Gorilla speak," "Recipe for a robot," and "Flight in a tunnel" provide facts about the science and technology, allowing readers to more fully understand the scientists' research goals. The authors (each experienced in writing about science for children) make complex topics of wildlife conservation, robotics, and biomechanics interesting and accessible—when a research question is posed, the reader will want to know the answer as well.

The books present compelling, attractive pictures of the scientists' personal and highly successful professional experiences. Because the whole of the scientists' lives are presented, young girls (and boys) will easily imagine translating their interests and skills (animals, sports, art, nature, school, etc.) into a scientific career. Vedder and Breazeal excelled at sports, and they each recalled how the team spirit and hard work inherent in athletics served them well when they needed to overcome obstacles and study hard during their scientific training. To determine how features such as bent knees and large webbed feet help tropical frogs fly, Koehl used her artistic talent to build models of frogs that allowed her to measure the physical effects of these design features.

In the inset on the ways gorillas communicate, Ebersole quotes Vedder's comment, "When gorillas 'sing', it can be like us humming in the shower. They're saying, 'hey, life is great!'" Vedder began her career in wildlife conservation at the Karisoke Research Center in Volcanoes National Park in Rwanda. She had resolved to help save the mountain gorillas from poaching and a dwindling habitat by learning more about the relationship between the gorillas and their forest environment. She tracked and observed gorillas from sunrise to sunset, recording diet and activities, learning that gorillas preferred to live in areas that contained a high variety of foods, and ultimately winning the confidence of several gorilla family groups. In one instance, Vedder and her husband Bill Weber hiked 16 hours, over the volcanoes spanning the border of Congo and Rwanda, carrying a young gorilla injured by a trapper's snare back to the research center. There, they cared for the frail ape, even performing cardiopulmonary resuscitation in an ultimately unsuccessful attempt to save its life so that it could be returned to the wild.

One lasting legacy of Vedder's work is the Mountain Gorilla Project, a conservation program that she and her husband proposed to the Rwandan government. When it was implemented in 1979, the project supported antipoaching teams, education about value of wildlife and natural resources, and gorilla tourism that brought jobs and foreign money to the country.

To study how people would interact with her emotionally responsive robot Kismet, Breazeal gave them the simple instruction "Speak to the robot." Breazeal began her career in robotics in Massachusetts Institute of Technology's Artificial Intelligence Laboratory, where she learned how to construct autonomous robots, machines that are programmed to function independ-

The reviewer is at the Arthritis and Immunology Research Program, MS 24, Oklahoma Medical Research Foundation, 825 N.E. 13th Street, Oklahoma City, OK 73104, USA. E-mail: Susan-Kovats@omrf.org

CREDIT: RYAN SNOOK

Published by AAAS

ently and make decisions on their own. For her own research, she constructed an emotionally expressive, socially intelligent robot that could respond to humans by conveying the spectrum of human emotions. She used 15 computers to control Kismet's expression of surprise, tiredness, happiness, sadness, or anger based on its fulfillment of "drives" such as the desire to interact with people, to play with brightly colored toys, or to rest after too much stimulation. Breazeal later collaborated with Stan Winston, the award-winning creator of robotic characters for Hollywood movies. Working together, their groups designed and constructed the autonomous robot Leonardo, a furry animal robot with the ability to see, hear, touch, and emotionally respond to people based on its prior interactions with them. Breazeal's ultimate goal is to develop robots that can cooperate with people as partners, by providing both service and companionship.

Discussing her discovery of how differences in the mechanical behavior of their tissues allow calm-water and wave-tossed anemones to feed by different techniques, Koehl declares, "I got really excited by my findings." "I like to know how organisms work." Koehl's research lies in biomechanics, a field in which the laws of physics are used to study how living organisms function in their environments. During her graduate studies in marine zoology, Koehl traveled to the Pacific coast of Washington state, where she studied the physical parameters that enable sea anemones to cling to the ocean bottom amid the crashing waves of a rugged coastline. Over the years, field investigations and experiments took her to remote coasts, where she often constructed her own equipment and waded into the turbulent tidal ecosystems to measure physical forces on the living "machines" she studied. Among her subsequent investigations that are discussed in the book is how specific design features (hairy or smooth) of noses or antennae are optimized to catch odor molecules swirling in currents. Koehl continues to devise creative techniques to understand how certain body designs allow organisms to eat and move in their particular environments.

One theme shared by the books, and the series as a whole, is reflected in advice that Koehl offers when talking to kids: "Every person has some gift or talent that can help unlock new answers about the world in which we live.… So don't be afraid of science, but do tackle it the way you do best."

References and Notes

[1] The scientific adventures of the 10 researchers are also showcased at a Web site that draws on the biographies, www.iwaswondering.org.

10.1126/science.1160700

CREDIT: RYAN SNOOK

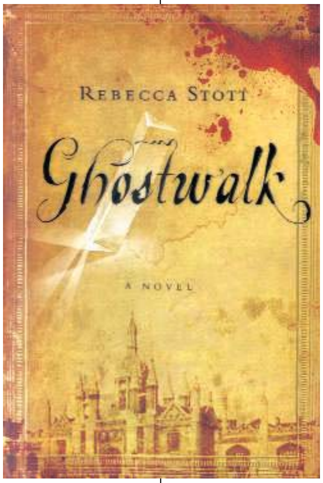
HISTORICAL FICTION

Newton's Ghosts

Jan Golinski

At the heart of the novel *Ghostwalk* is a mystery concerning the life of Isaac Newton: How did he come to be elected to a fellowship at Trinity College, Cambridge, in October 1667, at a time when very few people recognized his great mathematical genius? He needed the backing of powerful patrons, to be sure, but he also benefited from some recently created vacancies in the college fellowship. Three fellows of Trinity had died, apparently accidentally, by falling down stairs in the three previous years, and a fourth had been removed from the college on the grounds of insanity. Was Newton, then, the lucky beneficiary of the drunkenness and mental instability prevailing among Cambridge dons at the time? Or was something more sinister going on?

If historians have considered this question at all, they have probably concluded that it will never be possible to answer it with the evidence available. But Rebecca Stott, whose credentials as a scholar of the history of science are soundly demonstrated by her book *Darwin and the Barnacle* (*1*), decided to pursue it by writing a novel. She has produced an intelligent work of fiction, which also has things to say about the limits of scientific and historical knowledge. The book is a terrific summer read and great entertainment but also genuinely thought-provoking for scientists and historians.



Ghostwalk. *Rebecca Stott.* Spiegel and Grau, New York, 2007. 319 pp. $24.95. ISBN 9780385521062. Paper, $14.95. ISBN 9780385521079. Weidenfeld and Nicolson, London, £12.99. ISBN 9780297851363. Paper, £7.99. ISBN 9780753823576

Ghostwalk weaves together a plausible account of the 17th-century events surrounding Newton with a narrative set in contemporary Cambridge, the city being described with a precision that descends to the level of individual market stalls and street corners. The contemporary story also begins with a mysterious death, followed by a series of other disturbing and violent occurrences. These may be caused by ghostly revenants from the 17th century or by a shadowy terrorist group that is targeting scientists who experiment on animals. The connections between past and present emerge because the initial death is that of a scholar who was working on a book about Newton's researches on alchemy. One of the most original features of the novel is the inclusion of lengthy extracts from this imagined scholarly study, complete with genuine footnotes. Readers can learn a fair amount about Newton's experiments on light, his contacts with the glass-making trade, and his relations with alchemists and philosophers. One might conjecture that Stott's book began as an attempt to write this kind of scholarly study, and she later decided to build a novel around these passages.

Whether or not that is so, she has produced much more than a pastiche.

The reviewer is at the Department of History, University of New Hampshire, 20 Academic Way, Durham, NH 03824, USA. E-mail: jan.golinski@unh.edu

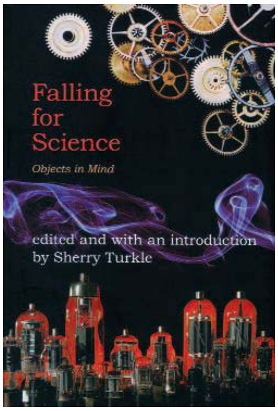
The contemporary narrative is partly a mystery and partly a love story. The narrator, Lydia Brooke, is smitten with Cameron Brown, a brilliant and glamorous neuroscientist with a penchant for extramarital affairs. Their romantic relationship is supposed to have ended before the novel begins, but is rekindled with a mixture of passion and self-loathing on Lydia's part, quite touchingly described. The story enters the realm of the occult when Lydia meets a medium who can call up spirits from the past, and the 17th-century mystery is eventually solved by a ghostly revelation of this kind.

Readers of a skeptical inclination may well find this denouement unsatisfying. But it seems that Stott wants to remind us that, in a way, we are still surrounded by ghosts even in the modern age. She frequently mentions how text messages and e-mails address us, emerging like the voices of disembodied spirits from the ether. The mysteries of quantum entanglement, which seems to involve instant communication over a long distance, also make an appearance. Perhaps Stott is also suggesting that a kind of attunement to the ghosts of the past can complement historical research? At the least, she has given her readers a wonderful example of how the creative imagination can take over when the interpretation of historical evidence has gone as far as it can.

References

1. R. Stott, *Darwin and the Barnacle* (Faber, London, 2003).

10.1126/science.1160708



SCIENCE COMMUNITY

First Adventures in Science

Falling for Science. Objects in Mind. *Sherry Turkle, Ed.* MIT Press, Cambridge, MA, 2008. 330 pp. $24.95, £16.95. ISBN 9780262201728.

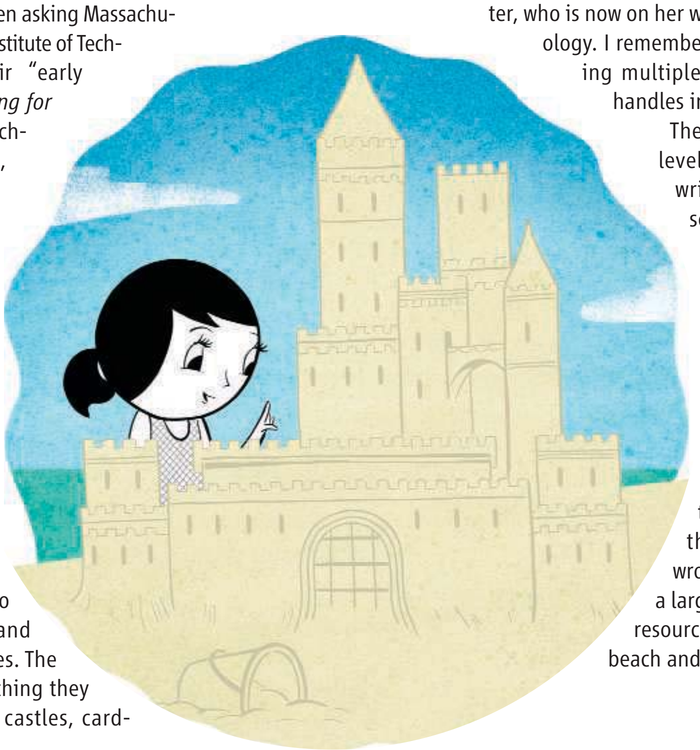
For more than 25 years Sherry Turkle has been asking Massachusetts Institute of Technology students to write about their "early moments of scientific curiosity." In *Falling for Science*, Turkle (a sociologist and psychologist in MIT's Program in Science, Technology, and Society) offers 51 of the students' stories about paths to science in which "imagination is sparked by an object." She groups these into sections focused around things seen, sensed, modeled, played with, built, sorted, and programmed. In addition, she provides essays from eight senior scientists, engineers, and designers ("mentors," now in their forties to their seventies) who recall similarly influential objects from their childhoods.

The book succeeds because the editor and contributors are not trying to force links between career outcomes and specific kinds of toys or early experiences. The writers share the experience that something they discovered as children—such as sand castles, cardboard boxes, fly fishing, or marbles—evoked in them an intensity of concentration that led to scientific habits of mind. In some cases, their early insights would later be rediscovered in their science classes. In others, the objects would lead to a way of exploring problems and thinking about the world that would last into their adult lives. Even the static electricity in a shirt can lead to a life-transforming "Aha!" moment.

Different personalities converged on scientific thinking in different ways. Sometimes, early curiosity was motivated by a desire for control—for example, those who as children could not look at a computer game without wanting to get into the computer and software and redesign them. Legos loomed large in "What we build," the one section of the book that may have gone on a bit too long. Here too, people took different approaches to these building blocks: one inventor as a child was most excited by following the instructions, whereas a budding information scientist was always trying to build something "better, bigger, cooler." Still others became obsessed with finding out how things worked. Neurobiologist (and MIT president) Susan Hockfield described her interest in taking things apart as a way to use structure to understand function.

For a number of the writers, science and art are not two separate disciplines, but have become fused in exhilarating adventures. Venus Paradise Pencil by Number Coloring Sets led cell biologist Donald Ingber to an understanding of "the power of the gestalt, that the whole is greater than the sum of its parts, and that the overall arrangement of the parts can be as important as the properties of these components." He has surely applied this understanding to his current studies of complex cellular systems. Timothy Bickmore (a computer scientist who develops relational agents) was a shy child who as a teenager became fascinated with lasers during a light show, learned how to build his own projectors, and put on shows. He considered the realization that "the pursuit of science from discovery to application could be an aesthetic experience" to have been crucial. And Sethby Cull made her way from analysis of the ingredients for baking the perfect chocolate meringue to studies of the interior of Mars.

The essays encourage reflection. Christine Alvarado (now teaching computer science at Harvey Mudd College) described how braiding her My Little Pony's hair had taught her about the mathematical concepts of division and recursion. Reading her story made me look back at my own daughter, who is now on her way to a career in mathematical sociology. I remember her sitting during car rides, working multiple colored strings attached to door handles into intricate knotting and braiding.



CREDIT: RYAN SNOOK

The collection can be read on several levels. The stories are uniformly well written and enjoyable. The editor's scholarly introduction and epilogue provide context for the essays as well as a thoughtful look at education in its broadest sense.

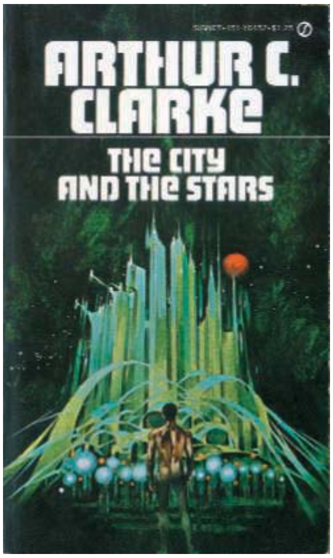
Falling for Science also evokes what may be one of the most important sensations to strive for in promoting creativity and certainly something to aim for during the summer: the sense of unlimited time. As Stephen Intille (now in the MIT Department of Architecture) wrote, "In school, there was never such a large block of uninterrupted time and a resource as boundless as a good stretch of beach and ocean."

–Barbara Jasny

10.1126/science.1161578

SCIENCE FICTION

Cities Not Built to Last



The City and the Stars. *Arthur C. Clarke.* Frederick Muller, London, 1956, 256 pp.

Sixty years ago, the visionary writer Arthur C. Clarke, who died 19 March (*1*), published his first novel, *Against the Fall of Night* (*2*), in the magazine *Startling Stories*. Not satisfied, he subsequently rewrote and expanded it as *The City and the Stars*. The earlier title, taken from a poem by A. E. Housman (*3*, *4*), more closely matches the novel's somber opening—Earth, a billion years hence: mountains ground to dust; all the land a desert wherein lies a lone city, Diaspar, refuge for the last remnant of a once Galaxy-spanning empire. The later version's title instead linked Diaspar to its destiny among the stars—a belief and desire that Clarke held also for humanity.

Clarke, one of the greats of 20th-century science fiction (along with such contemporaries as Isaac Asimov and Robert Heinlein), rehearsed in *The City and the Stars* themes that appeared in many of his subsequent books and stories. The plot, a simple and unerringly linear quest, is unremarkable even by the standards of the time, and the characters rarely rise above subservience to the story arc. What marks the book out are Clarke's sweeping vistas, grand ideas, and ultimately optimistic view of humankind's future in the cosmos.

The inhabitants of billion-year-old Diaspar—a breed of lotus-eating immortals "as carefully designed as [the city's] machines"—have ceded their Galactic empire to the "Invaders" and, on threat of their lives, have accepted confinement within Diaspar's city walls. The all-powerful Central Computer [a benevolent precursor to the more famous HAL from *2001: A Space Odyssey* (*5*)] runs the life of the city. Into this highly controlled environment comes Alvin, a "Unique," one of a vanishingly few individuals born anew.

Alvin's curiosity-fueled explorations, frowned on by his fellow immortals yet tacitly condoned by the Central Computer, reveal first a way out of their utopian prison and then its hidden sister settlement, the idyllic wooded villages of Lys. He eventually travels to the deserted ruins at the center of the Galaxy, where he finds the key to understanding the myths that imprison Earth and, ultimately, the true fate of the Galactic empire is revealed.

The inherent danger of scientific exploration—even Alvin worries about "the ruthless drive to satisfy his own curiosity"—forms the core of the somewhat-hurried final denouement. The myth of the Invaders is based upon the

BROWSINGS

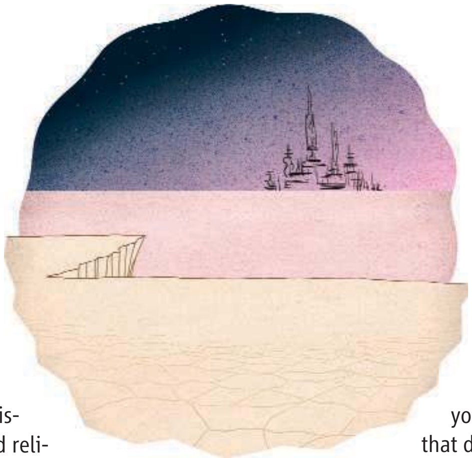
Disappearing World. 101 of the Earth's Most Extraordinary and Endangered Places. *Alonzo C. Addison.* Collins, New York, 2008. 272 pp. $34.95. ISBN 9780061434440. **Disappearing World.** The Earth's Most Extraordinary and Endangered Places. London, 2007. £25. ISBN 9780007261185.

Since 1972, the United Nations Educational, Scientific and Cultural Organization has offered the designation World Heritage to more than 850 cultural and natural sites deemed to be of "outstanding universal value." Although such recognition offers sites some protection, many remain threatened. Addison describes 101 of these treasured locations and highlights the one to several risks they face—which include conflict, theft, natural disasters, and climate change. His selection includes all 30 sites classified by UNESCO as in clear danger. Two sites protecting rainforest and waters come together at Iguazu Falls (below) on the Brazil-Argentina border; they are deemed "at risk" due to development and unsustainable tourism.

Into Thick Air. Biking to the Bellybutton of Six Continents. *Jim Malusa.* Sierra Club Books, San Francisco, 2008. 335 pp. Paper, $16.95. ISBN 9781578051410.

Instead of trying to bag the seven summits, the author chose solo descents to continental low spots: On month-long trips over six years, he pedaled—most of the way (weather and schedules occasionally led him to accept lifts)—from Darwin to Lake Eyre, Australia, Cairo to the Dead Sea, Moscow to the Caspian Sea, across the Andes to Salina Grande, through Djibouti to Lac Assal, and from his home in Tuscon to Death Valley. He recounts encounters with people, fauna, and flora along the way, but the strength of his narrative lies in his descriptions of the arid, often silent landscapes he so clearly loves.

CREDIT: DONALD NAUSBAUM/STONE/GETTY IMAGES



creation of a "pure mentality" that was "inspired and directed by Man" and that, in best monster tradition, promptly turns upon its creators and wreaks havoc across the Universe before it can be contained. Transcendence of the physical to something "other" also motivated Clarke's stories *Childhood's End* (*6*) and *2001*. There is no religious intent; Clarke's transcendence is simply a state of being beyond the limits of material experience. His (and our) incomprehension of this (engineered) transformation is a manifestation of his own third law: Any sufficiently advanced technology is indistinguishable from magic. Clarke's asides on organized religion show his disdain for the real thing clearly enough: it is a "meaningless morass" of ideas, suffers from "unbelievable arrogance," and engenders the "misplaced devotion" of its "deluded" adherents. The long-dead messiah of *The City and the Stars* fakes miracles, lies to his followers, and requires his lonely robot apostle to cover up for him.

Clarke's passion for exploration encompassed both outer space and the oceans. He was a keen scuba diver, and in 1954 he moved permanently to Ceylon (now Sri Lanka) so he could dive year-round. As a charming consequence, the alien creatures in the book all have a decidedly aquatic nature: for example, an intelligent colonial polyp that dissociates into its constituent cells when stressed, and huge gas-filled floating medusae harboring entire ecosystems in their trailing tendrils. These latter surely inspired those in his Nebula award–winning 1971 novella *A Meeting with Medusa* (*7*), and he developed the idea of large-scale underwater aquaculture in the novel *The Deep Range* (*8*).

For all its bravado and youthful vigor, I do not think the book (even in its rewritten form) has weathered as well as Clarke's reputation, and it has rightly been eclipsed by later works. Still, for those of us who read and fell in love with *The City and the Stars* in our youth—in my case, merely some 30 years ago—that doesn't matter one iota.

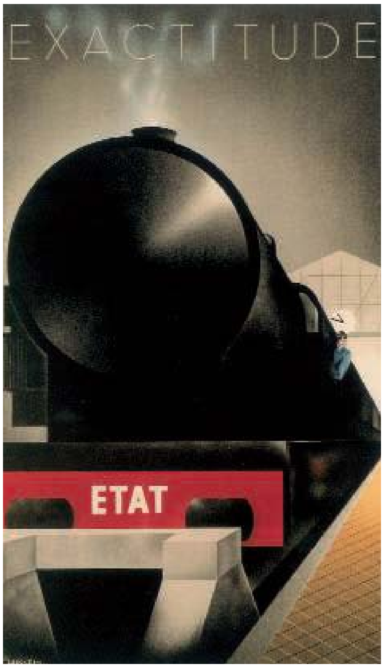
–Guy Riddihough

CREDIT: RYAN SNOOK

References and Notes

1. J. N. Pelton, J. Logsdon, *Science* **320**, 189 (2008).
2. A. C. Clarke, *Against the Fall of Night* (Gnome, New York, 1953).
3. A. E. Housman, "Smooth between sea and land," in *More Poems* (Cape, London, 1936); www.chiark.greenend.org.uk/~martinh/poems/complete_housman.html#MPxlv.
4. The title of this review is also taken from that poem.
5. A. C. Clarke, *2001: A Space Odyssey* (New American Library, New York, 1968).
6. A. C. Clarke, *Childhood's End* (Ballantine, New York, 1953).
7. A. C. Clarke, in *The Best of Arthur C. Clarke*, A. Wells, Ed. (Sidgwick and Jackson, London, 1973).
8. A. C. Clarke, *The Deep Range* (Frederick Muller, London, 1957).

10.1126/science.1161705

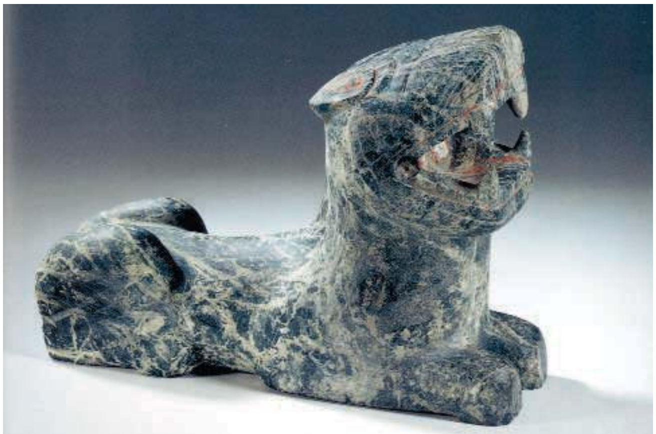


The Railway. Art in the Age of Steam. *Ian Kennedy and Julian Treuherz.* Yale University Press, New Haven, CT, 2008. 288 pp. $65, £35. ISBN 9780300138788. **Art in the Age of Steam.** Europe, America, and the Railway, 1830–1960. An exhibition at the Walker Art Gallery, Liverpool, through 10 August 2008; Nelson-Atkins Museum of Art, Kansas City, MO, 13 September 2008 to 18 January 2009.

The development of steam locomotives allowed people to travel faster than ever before—and changed their perception of space and time. Trains roared through landscapes, and scenery shot by passengers. This exhibition and catalog of paintings, prints (above, Pierre Fix-Masseau's color lithograph *Exactitude*, 1932), photographs, and posters explore how 19th- and 20th-century artists responded to the railway and its transformation of everyday life. The authors discuss a wide range of themes, including the railway in literature, human drama, the crossing of continents, Impressionism, and Modernism.

Discovery! Unearthing the New Treasures of Archaeology. *Brian M. Fagan, Ed.* Thames and Hudson, London, 2007. 256 pp. $40, C$50, £24.95. ISBN 9780500051498.

Many of the accounts in this global survey of remarkable findings from the past 15 years are provided by the discoverers themselves. The 62 succinct chapters (two to six pages) cover topics from early hominin fossils found in Africa to underwater studies of wrecks from the U.S. Civil War. They are grouped under seven broad categories (including tombs, graves, and mummies; ancient art; lost cities; and ritual and religion), but many could have fit in several of these sections. The final section comprises pieces on archaeological uses of DNA, the impact of past climate change, early alcoholic beverages, paleopathology, and several examples of ancient writing. Color photographs throughout the book depict sites, ruins, remains, and objects—such as this 30-cm-long stone carving of a crouching tiger (circa 1000 BCE) from Jinsha, Sichuan province, western China.



CREDITS: (LEFT) MODERNISM COLLECTION, GIFT OF NORWEST BANK MINNESOTA/MINNEAPOLIS MUSEUM OF ART/COURTESY YALE UNIV. PRESS
(RIGHT) CHENGDU MUSEUM/COURTESY THAMES AND HUDSON

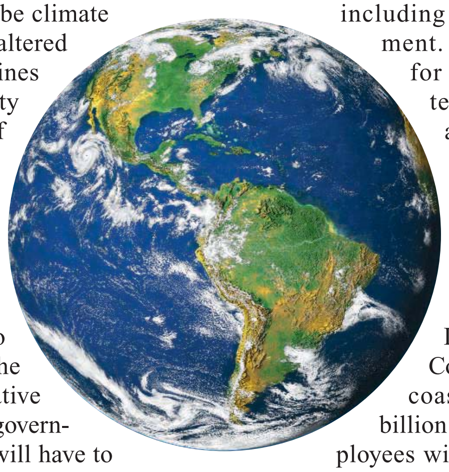
SCIENCE AND GOVERNMENT

An Earth Systems Science Agency

Addressing serious environmental and economic challenges in the United States will require organizational changes at the federal level.

Mark Schaefer,* D. James Baker, John H. Gibbons, Charles G. Groat, Donald Kennedy, Charles F. Kennel, David Rejeski

The United States faces unprecedented environmental and economic challenges in the decades ahead. Foremost among them will be climate change, sea-level rise, altered weather patterns, declines in freshwater availability and quality, and loss of biodiversity. Addressing these challenges will require well-conceived, science-based, simultaneous responses on multiple scales, from global and national, to regional and local. The executive and legislative branches of the federal government and of the states will have to transcend bureaucratic boundaries and become much more innovative in developing and implementing policy responses.



We strongly believe organizational changes must be made at the federal level to align our public institutional infrastructure to address these challenges. The most pressing organizational change that is required is the establishment of an independent Earth Systems Science Agency formed by merging the National Oceanic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS).

Current Organizational Structure

Two federal agencies, NOAA and USGS, have missions solely directed to the Earth sciences. NOAA's mission is directed primarily to the atmosphere and the oceans, including the coastal environment. USGS is responsible for freshwater and the terrestrial environment and has an extensive biological program. NOAA has a budget of nearly $4 billion and 12,000 employees, with research entities in the Washington, DC, area, in Boulder, Colorado, and along the coasts. USGS has a $1 billion budget and 8500 employees with administrative and research entities throughout the United States. Together, the two agencies are responsible for the major Earth science elements: air, land, water, and all living things.

The National Aeronautics and Space Administration (NASA) Earth science program is responsible for developing space-based Earth observing systems and performing associated research. NASA's Earth Science Program (*1*) budget is about $1.5 billion, with the bulk of its activities at the Goddard Space Flight Center in Maryland and Jet Propulsion Laboratory in California. Other important environmental research and development (R&D) activities take place in or through the National Science Foundation (NSF), the U.S. Environmental Protection Agency (EPA), the Department of Energy, the Department of Agriculture (USDA), the National Institutes of Health, and elsewhere.

Weaknesses in Federal Programs

Federal environmental research, development, and monitoring activities are not presently structured to respond to the challenges of today and tomorrow. To illustrate this, we point to Earth observation systems, one of several compelling examples.

Robust Earth-observing systems are critical to meeting national and international needs. Yet these systems have not kept pace with increasing demands of the public and private sectors for comprehensive, high-quality information on the changing global environment. At a time when federal Earth-observing systems should have been ramping up, priorities have shifted to manned missions to the Moon and Mars. A recent study by the National Research Council found that NASA's Earth science budget had declined 30% since 2000 (*2*). The scientific importance and societal value of remote sensing systems has not been communicated effectively to the public and Congress; hence, there is little awareness of the shortfalls in our Earth-observing systems—and no driving force to address them. Yet these systems are critical to public safety, natural disaster response, and efficient transportation and they fuel multibillion-dollar industries.

The synergies among our research and monitoring programs, both space- and ground-based, are not being exploited effectively because they are not planned and implemented in an integrated fashion. Our problems include inadequate organizational structure, ineffective interagency collaboration, declines in funding, and blurred authority for program planning and implementation.

Earth observation programs cut across NOAA, the USGS, NASA, and other agencies including the NSF, EPA, and USDA. The total budget for federal environmental R&D programs is nearly $8 billion. Despite the magnitude of the nation's environmental challenges, funding trends for federal programs have been downward or at best flat in recent years. The Administration's FY 2009 request for R&D Earth science funding for USGS and NOAA, as well as at EPA and USDA, includes further declines.

A Proposed Earth Systems Science Agency

We propose that an Earth Systems Science Agency (ESSA) be formed by combining NOAA and USGS and by building a strong policy, administrative, and collaborative research bridge to NASA's Earth sciences program. The agency should focus on research, monitoring, communication, and the advancement of applications, particularly decision support systems that inform policy-making and guide implementation. It should not have direct regulatory responsibilities. Although some NASA analysis and applications elements

Each of the coauthors has held senior Earth and environmental science positions in the federal government, including M. Schaefer: Deputy Assistant Secretary of the Interior, Acting Director of the U.S. Geological Survey; D. J. Baker: Administrator, National Oceanic and Atmospheric Association; J. H. Gibbons: Director, White House Office of Science and Technology Policy, Science Adviser to the President; C. G. Groat: Director, U.S. Geological Survey; D. Kennedy: Commissioner, Food and Drug Administration; C. F. Kennel: Associate Administrator, National Aeronautics and Space Administration, Director of Mission to Planet Earth; D. Rejeski: Served at the Office of Science and Technology Policy and Council on Environmental Quality from 1994 to 2000.

Author contact e-mail: djamesbaker@comcast.net, jackgibbons@hughes.net, cgroat@mail.utexas.edu, kennedyd@stanford.edu, ckennel@ucsd.edu, david.rejeski@wilsoncenter.org

*Author for correspondence. E-mail: markschaefer24@msn.com

CREDIT: RETO STOCKLI, ALAN NELSON, FRITZ HASLER/NASA

could be incorporated into ESSA, most of NASA's Earth sciences research and observation program should remain in its present organizational location to allow it to continue to capitalize on NASA space technology. NASA has worked effectively with NOAA for decades, and it could work equally well with the new agency. However, NASA should be directed both to restore Earth systems science as a prime agency mission and to work collaboratively with ESSA. NASA's space technology is key to the success of ESSA. We believe NASA's satellite systems need to focus first and foremost on planet Earth, the planet that sustains human life.

ESSA should be an independent federal agency, which would allow it to support all federal departments and agencies and would give its director direct access to the Congress and the Executive Office of the President, including the Office of Science and Technology Policy and the Office of Management and Budget.

To be effective, ESSA must coordinate its research and development activities with those of the NSF, Department of Energy and its national laboratories, EPA, National Institute of Environmental Health Sciences, Department of the Interior, Department of Commerce, USDA, and other agencies. The White House Office of Science and Technology Policy, the Office of Management and Budget, and the National Science and Technology Council will need to foster interagency collaboration and to ensure adequate funding of Earth systems science programs. Also, mechanisms to link ESSA's activities with state agencies will be needed.

The core mission of ESSA should be to conduct and sponsor research, development, monitoring, educational, and communications activities in Earth systems science. Its portfolio should include ocean, atmospheric, terrestrial, cryosphere, freshwater, and ecological processes and the interactions among them. It should develop and communicate comprehensive information on Earth processes, including natural disasters and extreme weather events. It should generate information critical to the sustainable use of water, mineral, biomass, wind, and other resources. Also, it should provide information on the state and quality of freshwater, estuarine, and marine biological resources and nonrenewable materials resources to guide commercial and conservation activities.

The private sector already relies heavily on data and information products from NOAA and USGS. Information on weather, natural disasters, water quality, geology, geography, fisheries, and other biological resources fuels a large, multibillion-dollar private sector enterprise, as well as directly supports individuals and nonprofit organizations. A new generation of integrated products and services available under ESSA would foster private sector innovation and spur economic development.

ESSA's success will depend largely on its ability to generate and communicate reliable scientific information to the public and private sectors. This will require effective advisory bodies, internal and external peer review mechanisms, and communications and outreach capabilities.

Building on the excellent base already in place, ESSA can become a major home of world-class Earth sciences research, an institution that engages the best Earth and environmental scientists in the nation, and a focal point for collaboration with outstanding researchers internationally. Through its reputation and programs, the agency would attract a new generation of scientists and engineers.

PHOTO CREDITS: (TOP LEFT) JUPITER IMAGES; (TOP RIGHT) © CORBIS. ALL RIGHTS RESERVED; (BOTTOM LEFT) USGS; (BOTTOM RIGHT) JUPITER IMAGES

No less than 25% of ESSA's budget should be devoted to grants, contracts, and cooperative agreements with academic and nonprofit institutions. ESSA should coordinate its extramural activities with the grant-making efforts of the NSF.

To be successful, the new agency will need to build on academia's basic research accomplishments, as well as its specialized organizational and technological capabilities. This includes high-performance computing, modeling, visualization, and monitoring expertise and technologies. In addition, we believe that a proportion of the new agency's R&D funding should be set aside and managed to target opportunities that cut across disciplinary boundaries and foster breakthrough technologies, along the lines of the Defense Advanced Research Projects Agency.

ESSA must be organized with the guidance and support of Congress. Committee and subcommittee responsibilities should be aligned to further congressional oversight responsibilities. Champions on Capitol Hill have been critical to the success of other federal agencies. ESSA will need congressional champions as well.

Creating new organizational entities within our federal government is rare, but not unprecedented. Between 1936 and 1973, six commissions were created to explore the reorganization of the executive branch. One of these entities, the Ash Council, laid the groundwork for the creation of the Environmental Protection Agency in 1970, which integrated a half-dozen functions from agencies such as Interior; Health, Education and Welfare; and USDA. As with EPA, new agencies often arise in response to a sudden or compelling national need.

We call on the next U.S. President and Congress to act quickly to realign federal Earth sciences R&D programs, provide them adequate funding, and ensure that they are closely linked to the wealth of talent in the nation's academic institutions. Convening a commission similar to the Ash Council would be an effective way to define a path forward.

References

1. *NASA Earth Science Applications Plan* (Office of Earth Science, NASA, Washington, DC, 2004).
2. National Research Council, *A Review Assessment of NASA's Applied Sciences Program* (Committee on Extending Observations and Research Results to Practical Applications: A Review of NASA's Approach, National Academies Press, Washington, DC, 2007).

10.1126/science.1160192

NEUROSCIENCE

The Scale of Experience

Specific cells in the hippocampus allow the rat brain to track spatial location at different scales.

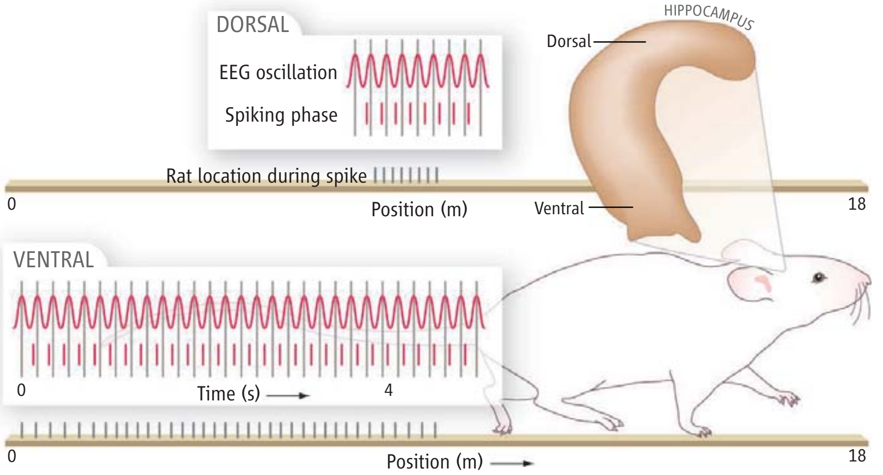
Michael E. Hasselmo

Most people would not equate remembering their way around the neighborhood with remembering their way around the kitchen, but the same neural mechanisms may be involved in navigating on both scales. On page 140 in this issue, Kjelstrup *et al*. (*1*) show that neurons at different anatomical positions along the length of the rat hippocampus may represent location along a continuum of spatial scales.

Single neurons in the rat hippocampus known as place cells fire selectively when the rat moves through specific locations (*2*). The area where a particular place cell fires is known as its place field. Experimental studies of place cells usually focus on the dorsal hippocampus, where most cells show relatively small place fields (<50 cm in diameter). However, behavioral data indicated that selective ventral hippocampus lesions do not impair spatial memory on a small scale (*3*), but do impair contextual fear conditioning (*4*), motivating Kjelstrup *et al*. to test whether the ventral hippocampus codes space on a larger scale.

Ventral hippocampal neurons are difficult to target and rarely recorded. Two studies showed differences in place field size between dorsal and intermediate hippocampus (*5*, *6*), but another reported that dorsal and ventral cells have more similar characteristics (*7*) possibly due to the use of a small recording environment, as the measured place field size increases with environment size (*2*, *8*).

One innovation of the current study is the use of an exceptionally large recording environment. Most place cell studies use environments about 1 m across (a short sprint for a rat), because in larger environments it is difficult to track the location of rats and avoid impediments to their movement. Overcoming these technical issues, researchers installed an 18-m track through hallways in the Moser laboratory, allowing them to quantify activity on large spatial scales (a longer-distance run for a rat). Ventral hippocampal neurons showed dramatically large firing fields covering distances greater than 10 m, whereas dorsal neurons fired over a mean length of 98 cm (see the figure). These data are consistent with changes in the spatial scale of grid cells in medial entorhinal cortex (*9*). The medial entorhinal cortex provides a major input to the hippocampus, and grid cells exhibit a repeating pattern of firing fields that could provide a basis for driving place cell firing. Grid cells also show progressively increasing spatial scale along the dorsal-to-ventral axis of entorhinal cortex (*9*), culminating in very large fields in ventral regions (*10*).



Ventral hippocampal neurons fire with larger place fields than dorsal cells as a rat runs on a track. Oscillatory traces show how an interference model of grid cells (*11*) could account for the difference in spatial scale and time course of phase precession if running causes smaller changes in the oscillatory frequency of neural activity in ventral as compared with dorsal cells (*11*–*13*).

CREDIT: P. HUEY/SCIENCE

How does the brain represent such different spatial scales? The time course of neural activity in the large fields exceeds the time constants of most neuronal properties, though persistent firing mechanisms or recurrent excitation may contribute. In physics, interference phenomena are used for measurements at multiple scales, from the molecular to the astronomical. The brain may similarly use interference phenomena based on oscillations in the activity of neurons to encode distance.

A model of grid cells based on interference of oscillations in membrane potential or neuron firing (*11*) can account for the dorsal-ventral increase in the spatial scale of grid fields (*11*, *12*). The model predicted a difference in the intrinsic frequency of neurons along the dorsal-to-ventral axis of entorhinal cortex that was supported by intracellular recording of subthreshold membrane potential oscillations in entorhinal neurons in slices (*13*). Model simulations (*12*) can replicate differences in grid scale, including the large grid fields found in ventral entorhinal cortex (*10*). The model (*11*) generates a change in the phase of grid cell firing relative to theta frequency oscillations in the local field potential that is proportional to firing field size (see the figure), potentially accounting for place cell precession on many scales (*1*, *6*, *11*). Notably, data on the period of membrane potential oscillations show smaller variance between neurons in dorsal versus ventral entorhinal cortex (*13*), resembling the smaller variance of place field size in dorsal versus ventral hippocampus (*1*). The model (*11*) also predicted the smaller differences seen between the intrinsic firing frequency of neurons and network theta rhythm in more ventral cells (*1*).

On a behavioral level, many studies have focused on a difference in behavioral function between the dorsal and ventral hippocampus. Dorsal but not ventral hippocampal lesions impair spatial memory performance (*3*). Ventral hippocampal lesions can alter behavior with an affective component, such as defecation and entry to open areas (*3*), or context-dependent fear conditioning (*4*).

The different scale of place field firing (*1*) could explain some functional differences between dorsal and ventral hippocampus. Learning the location of a small platform in a spatial memory task may require the high resolution of dorsal place fields (*3*), whereas the large spatial scale of ventral place field activity could allow association of a particular room with footshock (*4*). Ventral neurons fire almost

Center for Memory and Brain, Department of Psychology and Program in Neuroscience, Boston University, Boston, MA 02215, USA. E-mail: hasselmo@bu.edu

Published by AAAS

everywhere in an environment in one room, and nowhere in an identical environment in another room (*1*). Effects interpreted as context may arise from representing experience at a large scale. Learning to avoid aversive stimuli may require a larger scale than required for other stimuli, resulting in an evolutionary advantage for stronger connectivity from ventral hippocampus to structures involved in fear responses such as the amygdala and hypothalamus. Even our daily experience suggests a difference in scale for fear versus object location. You may feel sweaty palms and pounding heartbeat in an alley in a bad part of town, but your heart rate does not change as you walk past the gas stove or the garbage disposal in the kitchen (potentially more dangerous locations, but on a smaller scale). Hippocampal neurons might also reflect the scale of other dimensions of memory (*14*). For instance, the ventral hippocampus might be involved in associations on a larger temporal scale (*15*).

These place field data suggest that behavioral differences between dorsal and ventral hippocampus may reflect different scales of experience. The effect of lesions on different behavioral scales could be tested systematically. The largest scale discussed here resembles the scale of rat territory (*1*), but species such as humans might have cells coding even larger scales, such as segments of one's morning commute.

References and Notes

1. K. B. Kjelstrup *et al.*, *Science* **321**, 140 (2008).
2. J. O'Keefe, N. Burgess, *Nature* **381**, 425 (1996).
3. K. G. Kjelstrup *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10825 (2002).
4. M. A. Richmond *et al.*, *Behav. Neurosci.* **113**, 1189 (1999).
5. M. W. Jung, S. I. Wiener, B. L. McNaughton, *J. Neurosci.* **14**, 7347 (1994).
6. A. P. Maurer, S. R. Vanrhoads, G. R. Sutherland, P. Lipa, B. L. McNaughton, *Hippocampus* **15**, 841 (2005).
7. B. Poucet, C. Thinus-Blanc, R. U. Muller, *Neuroreport* **5**, 2045 (1994).
8. R. U. Muller, J. L. Kubie, J. B. Ranck Jr., *J. Neurosci.* **7**, 1935 (1987).
9. T. Hafting, M. Fyhn, S. Molden, M. B. Moser, E. I. Moser, *Nature* **436**, 801 (2005).
10. T. Solstad *et al.*, *Soc. Neurosci. Abstr.* **33**, 93.2 (2007).
11. N. Burgess, C. Barry, J. O'Keefe, *Hippocampus* **17**, 801 (2007).
12. M. E. Hasselmo, L. M. Giocomo, E. A. Zilli, *Hippocampus* **17**, 1252 (2007).
13. L. M. Giocomo, E. A. Zilli, E. Fransen, M. E. Hasselmo, *Science* **315**, 1719 (2007).
14. H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, H. Tanila, *Neuron* **23**, 209 (1999).
15. T. Yoon, T. Otto, *Neurobiol. Learn. Mem.* **87**, 464 (2007).
16. Supported by the National Institute of Mental Health (grants MH71702, MH60013, MH61492, and MH60450), NSF Science of Learning Centers program (grant 0354378), and National Institute on Drug Abuse (grant DA16454).

10.1126/science.1160121

SOCIOLOGY

Indirect Social Influence

To what extent are our decision-making and learning processes influenced indirectly by others?

Jerker Denrell

To what extent are the opinions you hold simply a reflection of the opinions of those you associate with? Most people like to think that their opinions are based on their own deliberations. Of course, there are exceptions. You may take into account the opinions of others if you believe they are better informed. You may even conform to the majority opinion in order to avoid being seen as deviant (*1*, *2*). Studies of how norms and beliefs vary between groups, and how they are transmitted from peers or parents, testify to the importance of such social influence (*3*).

Explanations of social influence usually focus on why people are persuaded by or conform to the opinions of others (*4*). Although important, this research has neglected the role of information collection in belief formation and how biased beliefs, as well as social influence, can emerge from biased search processes (*5*).

THEY'RE A SAMPLING OF SOCIAL EQUALS WHO INDIRECTLY AFFECT MY OPINION...
SALE!

CREDIT: JOE SUTLIFF

For example, suppose you are deciding which of two cars to buy. If your neighbor buys one of the cars, you can observe it more closely and will thus learn more about its attributes. This opportunity to observe the car can bias your decision toward buying the same car, even if you do not care about whether you have the same car as your neighbor. This is especially true if acquiring information about cars other than your neighbor's is costly (*6*). If the information you learn about your neighbor's car is strongly positive, it makes sense to buy this car and discontinue the search. In this case you will not find out whether the other car is superior. If the information you learn is not very positive, however, it then makes sense to examine the other car. Only in this case will you find out how the two cars compare. Because the comparison process is asymmetric, you are overall more likely to buy the same car as your neighbor even if the information you learn is equally likely to be positive or negative.

The attitudes and behavior of others can also influence our learning processes by leading us to revisit objects and events that we had previously avoided because of poor past experiences (*7*). Suppose Bob likes a restaurant while Alice does not. By herself Alice might not visit the restaurant again,

Stanford Graduate School of Business, Stanford University, Stanford, CA 94305, USA. E-mail: denrell@gsb.stanford.edu

Published by AAAS

and her attitude would remain negative. But Alice might join Bob if he wants to go to the restaurant. By visiting the restaurant again, Alice gets a chance to change her opinion. Alice's attitude will depend on Bob's, but only because he influenced the probability of her revisiting the restaurant.

Finally, the number of your friends who engage in some activity can also influence your estimate of the value of this activity. If you have many friends who start firms, for example, your estimate of the chances of success will be based on a large sample size. A large sample size may lead you to have a higher estimate of the success rate than you would if the sample size were small. Experiments show that a large sample size leads to a more optimistic view when the outcome distribution is skewed (*8*). If only 10% succeed, you may only observe failures in a small sample, and will then underestimate the success rate.

These mechanisms produce behavior that looks like conformity: You are more likely to evaluate an activity positively if others do so. But in these examples your attitude is not directly influenced by hearing about the attitudes of others. Your attitude is only indirectly influenced by others because their behavior exposes you to additional samples of the activity.

Such indirect mechanisms of social influence are important, because even individuals who try to be impartial and make the best decision given the available information may fail to recognize that the available information is influenced by others (*9*). For example, a manager who tries to avoid discrimination may nevertheless come to believe that individuals who belong to the same social networks as the manager does are superior to those the manager seldom interacts with and has less information about. To learn more about these mechanisms, we need to broaden studies of social influence and belief formation to include the phases of learning and information collection that precede decision-making and judgment (*10*).

References and Notes

1. S. E. Asch, *Sci. Am.* **193**, 31 (November, 1955).
2. M. Deutsch, H. Gerard, *J. Abnorm. Soc. Psychol.* **51**, 629 (1955).
3. P. J. Richerson, R. Boyd, *Not by Genes Alone: How Culture Transformed Human Evolution* (Univ. of Chicago Press, Chicago, 2005).
4. R. B. Cialdini, N. J. Goldstein, *Annu. Rev. Psychol.* **55**, 591 (2004).
5. Y. Trope, A. Liberman, in *Social Psychology: Handbook of Basic Principles*, E. T. Higgins, A. W. Kruglanski, Eds. (Guilford, New York, 1996), pp. 239–270.
6. N. V. Moshkin, R. Shachar, *Market. Sci.* **21**, 435 (2002).
7. J. Denrell, G. Le Mens, *Psychol. Rev.* **114**, 398 (2007).
8. R. Hertwig *et al.*, *Psychol. Sci.* **15**, 534 (2004).
9. J. Denrell, *Psychol. Rev.* **112**, 951 (2005).
10. For recent research on the effect of sampling on judgment, see K. Fiedler, P. Juslin, Eds., *Information Sampling and Adaptive Cognition* (Cambridge Univ. Press, Cambridge, 2006).
11. I thank J. March for comments.

10.1126/science.1157667

NEUROSCIENCE

Transient Dynamics for Neural Processing

A computational view of how perception and cognition can be modeled as dynamic patterns of transient activity within neural networks.

Misha Rabinovich,[1] Ramon Huerta,[1] Gilles Laurent[2]

Neural networks are complicated dynamical entities, whose properties are understood only in the simplest cases. When the complex biophysical properties of neurons and their connections (synapses) are combined with realistic connectivity rules and scales, network dynamics are usually difficult to predict. Yet, experimental neuroscience is often based on the implicit premise that the neural mechanisms underlying sensation, perception, and cognition are well approximated by steady-state measurements (of neuron activity) or by models in which the behavior of the network is simple (steady state or periodic). Transient states—ones in which no stable equilibrium is reached—may sometimes better describe neural network behavior. An intuition for such properties arises from mathematical and computational modeling of some appropriately simple experimental systems.

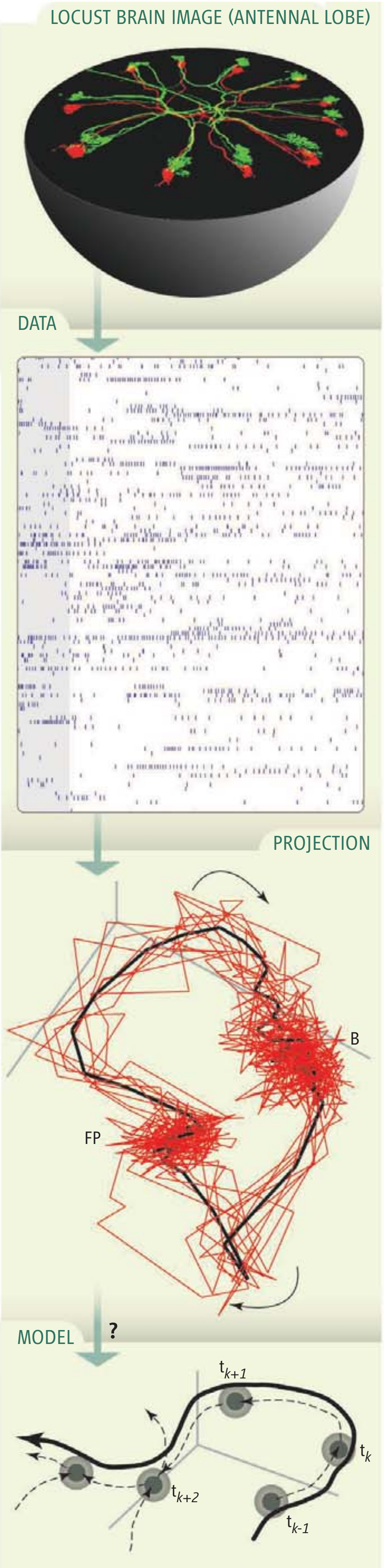
Computing with "attractors" is a concept familiar to the neural networks community. Upon some input signal, a model neural network will gradually change its pattern of activated nodes (neurons) until it settles into one pattern—an attractor state. Thus, the input—a voice, an odor, or something more abstract—is associated with properties of the entire network in a particular attractor state. Such patterns of neural activity might be established, learned, and recalled during perception, memorization, and retrieval, respectively.

Two ideas define the range of possible dynamics expressed by neural networks. The simplest emphasizes stable attractors (*1*), with memories as possible cognitive equivalents. The other, less intuitive, idea emphasizes nonclassical, transient dynamics as in "liquid-state machines" (*2*). Liquid-state machines are networks in which computation is carried out over time without any need for a classical attractor state. Because neural phenomena often occur on very short time scales, classical attractor states—fixed points or limit cycles—cannot be realistically reached. Indeed, behavioral and neurophysiological experiments reveal the existence and functional relevance of dynamics that, while deterministic, do not require waiting to reach classical attractor states (*3*–*6*). Also, the conditions required to achieve such attractors in artificial neural networks are often implausible for known biological circuits. Finally, fixed-point attractor dynamics, despite their name, express no useful dynamics; only the state the network settles into, given by its initial conditions (and characterized mathematically by, for example, a minimum in an energy function), matters, not the path taken to reach that state.

An alternative theoretical framework may explain some forms of neural network dynamics that are consistent both with experiments and with transient dynamics. In this framework, transient dynamics have two main features. First, although they cannot be described by classical attractor dynamics, they are resistant to noise, and reliable even in the face of small variations in initial conditions; the succession of states visited by the system (its trajectory, or transient) is thus stable. Second, the transients are input-specific, and thus contain information about what caused them in the first place. Notably, systems with few degrees of freedom do not, as a rule, express transient dynamics with such properties. Therefore, they are not good models for developing the kind of intuition required here. Nevertheless, stable transient dynamics can possibly be

[1] Institute for Nonlinear Science, University of California at San Diego, La Jolla, CA 92093, USA. [2] Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: laurentg@caltech.edu

Published by AAAS



CREDITS: (TOP) SARAH FARIVAR. (MIDDLE) UPPER: OFER MAZOR AND STIJN CASSENAER

Brain computation with transients. A model of how neural networks in the locust antennal lobe process information. (**Top**) Schematic of an antennal lobe sectioned through its equatorial plane. Principal neurons are labeled red and green. (**Middle**) Upper: Single-trial responses of 110 locust antennal-lobe principal neurons to one odor can be recorded (gray bar, 1 s). Lower: Projections of principal-neuron trajectories, representing the succession of states visited by this neural network in response to one odor. Red lines, individual trials; black line, average of 10 trials. B, baseline state; FP, fixed point, reached after 1.5 s. [Reprinted from (*10*) with permission from Elsevier] (**Bottom**) Putative dynamical model of transients: a set of dissipative saddles (dark circle), sequentially connected by unstable separatrices (dashed lines). A single trajectory (continuous line) connects the neighborhoods of saddles in a heteroclinic channel. $t_{k+1} - t_k$ is the characteristic interval needed for the transition between saddle points.

understood from within the existing framework of nonlinear dynamical systems.

Experimental observations in the olfactory systems of locust (*7*) and zebrafish (*8*) support such an alternative framework. Odors generate distributed (in time and space), odor- and concentration-specific patterns of activity in principal neurons. Hence, odor representations can be described as successions of states, or trajectories, that each correspond to one stimulus and one concentration (*9*). Only when a stimulus is sustained does its corresponding trajectory reach a stable fixed-point attractor (*10*). However, stable transients are observed whether a stimulus is sustained or not—that is, even when a stimulus is sufficiently short-lived that no fixed-point attractor state is reached. When the responses to several stimuli are compared, the distances between the trajectories corresponding to each stimulus are greatest during the transients, not between the fixed points (*10*). Because transients and fixed points represent states of neuronal populations, and because these states are themselves read out or "decoded" by yet other neuronal populations, stimulus identification by such decoders should be more reliable with transient than with fixed-point states. This conclusion is supported by the observation that a population of neurons that receives signals from the principal neurons responds mostly during transients, when separation between inputs is optimized. In response to these observations, a theoretical framework needs to explain the system's sensitivity to incoming signals, its stability against noise (external noise and intrinsic pulsations of the system), and its minimal dependence on the initial conditions (reproducibility).

To understand such transient dynamics, a mathematical image is needed that is consistent with existing results, and its underlying model(s) must be used to generate testable predictions. One possible image is a stable heteroclinic channel (*11*, *12*) (see the figure). A stable heteroclinic channel is defined by a sequence of successive metastable ("saddle") states. Under the proper conditions, all the trajectories in the neighborhood of these saddle points remain in the channel, ensuring robustness and reproducibility in a wide range of control parameters. Such dynamical objects are rare in low-dimensional systems, but common in complex ones. A possible underlying model is a generalized Lotka-Volterra equation (see supporting online material), which expresses and predicts the fate of an ongoing competition between *n* interactive elements. When *n* is small (for example, two species competing for the same food source, or predator-prey interactions), limit cycles are often seen, consistent with observations (*13*). When *n* is large, the state portrait of the system often contains a heteroclinic sequence linking saddle points. These saddles can be pictured as successive and temporary winners in a nonending competitive game. In neural systems, because a representative model must produce sequences of connected neuron population states (the saddle points), neural connectivity must be asymmetric, as determined by theoretical examination of a basic "coarse grain" model (*12*). Although many connection statistics probably work for stable heteroclinic-type dynamics, it is likely that connectivity within biological networks is, to some extent at least, the result of optimization by evolution and synaptic plasticity.

What are the conditions necessary for transient stability? Consider a three-dimensional autonomous inhibitory circuit with asymmetric connections. Such a system displays stable, sequential, and cyclic activation of its components, the simplest variant of a "winner-less" competition (*11*). High-dimensional systems with asymmetric connections can generate structurally stable sequences—transients, each shaped by one input (*14*). A stable heteroclinic channel is the dynamical image of this behavior (see the figure).

Asymmetric inhibitory connectivity also helps to solve the apparent paradox that sensitivity and reliability in a network can coexist (*12*, *14*, *15*). To be reliable, a system must be both sensitive to the input and insensitive to perturbations and initial conditions. To solve this paradox, one must realize that the neurons participating in a stable heteroclinic channel are assigned by the stimulus, by virtue of their direct and/or indirect input from the neurons activated by that stimulus. The joint action of the external input and a stimulus-dependent connectivity matrix defines the stimulus-specific heteroclinic channel. In addition, asymmetric inhibition coordinates the sequential activity of the neurons and keeps a heteroclinic channel stable.

The idea behind a liquid-state machine is based on the proposals that the cerebral cortex is a nonequilibrium system and that brain computations can be thought of as unique patterns of transient activity, controlled by incoming input (*2*). The results of these computations must be reproducible, robust against noise, and easily decoded. Because a stable heteroclinic channel is possibly the only dynamical object that satisfies all required conditions, it is plausible that "liquid-state machines" are dynamical systems with stable heteroclinic channels, based on the principle of winner-less competition. Thus, using asymmetric inhibition appropriately, the space of possible states of large neural systems can be restricted to connected saddle points, forming stable heteroclinic channels. These channels can be thought of as underlying reliable transient brain dynamics. It will be interesting to see if extensions of these ideas can apply to large neural circuits, and to the perceptual and cognitive functions that they subserve.

References

1. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A* **79**, 2554 (1982).
2. W. Maass, T. Natschlager, H. Markram, *Neural Comput.* **14**, 2531 (2003).
3. N. Uchida, Z. F. Mainen, *Nat. Neurosci.* **6**, 1224 (2003).
4. L. Lin, R. Osan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102,** 6125 (2005).
5. E. H. Baeg *et al.*, *Neuron* **40**, 177 (2003).
6. M. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 18772 (2007).
7. G. Laurent, H. Davidowitz, *Science* **265**, 1872 (1994).
8. R. Friedrich, G. Laurent, *Science* **291***,* 889 (2001).
9. M. Stopfer, V. Jayaraman, G. Laurent, *Neuron* **39**, 991 (2003).
10. O. Mazor, G. Laurent, *Neuron* **48**, 661 (2005).
11. M. Rabinovich *et al.*, *Phys. Rev. Lett.* **87**, 068102 (2001).
12. R. Huerta, M. Rabinovich, *Phys. Rev. Lett.* **93**, 238104 (2004).
13. T. Lindström, *J. Math. Biol.* **31**, 6 (1993).
14. T. Nowotny, M. Rabinovich, *Phys. Rev. Lett.* **98**, 128106 (2006).
15. M. Rabinovich, R. Huerta, P. Varona, V. Afraimovich, *PLoS Comput. Biol.* 4, e1000072 (2008).

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/48/5885/DC1
SOM Text

10.1126/science.1155564

MATERIALS SCIENCE

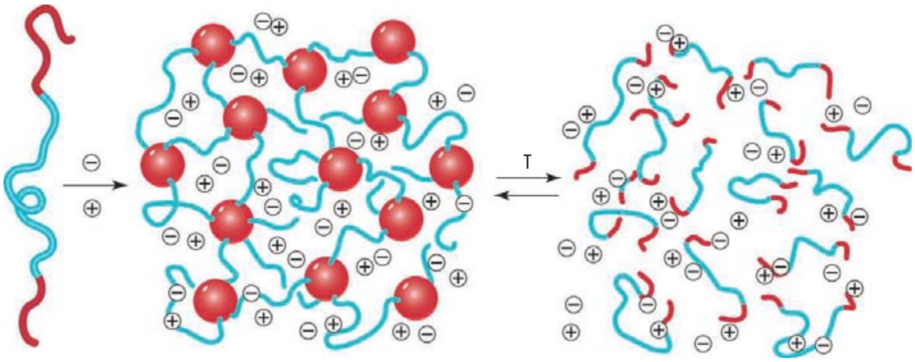
A Unique Platform for Materials Design

By combining ionic liquids with block polymers, a virtually unlimited range of composite materials can be prepared.

Timothy P. Lodge

Ionic liquids have generated excitement over the past decade as solvents for chemical transformation. For example, they have enabled dissolution of biopolymers such as cellulose. Less noted, but no less important, is their viability as materials. Their negligible vapor pressure and exceptional chemical, electrochemical, and thermal stability are attributes traditionally associated with solids, yet ionic liquids retain the high ionic conductivity and rapid response to external stimuli that characterize liquids. The two key properties of solids that ionic liquids lack—mechanical integrity and persistent structure—can be supplied by mixing with suitable polymeric components such as block polymers.

Ionic liquids are mixtures of low-molar-mass anions and cations. Thousands of ions can be combined in millions of different ways to optimize properties. Block polymers are macromolecules containing two or more different repeat units (monomers) that are covalently linked in contiguous sequences (blocks). The immiscibility of the different blocks drives segregation on the length scale of the blocks, leading to a variety of self-assembled nanostructures. Monomers, block lengths, and block sequence can be selected to achieve a desired set of attributes.



Reversible transition. An ABA triblock copolymer with soluble B block (blue) and insoluble A blocks (red) (**left**) self-assembles in the presence of an ionic liquid (+ and – symbols) to form an ion gel (**center**). Suitable choice of A block enables thermoreversible gelation when the A blocks become soluble at a higher temperature (**right**).

CREDIT: ADAPTED FROM TIMOTHY P. LODGE

Because both ionic liquids and block polymers offer virtually unlimited tunability, their composites are a unique platform for designing materials (*1*). Possible applications include membranes for fuel cells and gas separations, ionic conduction media for electrochemical sensors and batteries, electroresponsive gels for actuators and artificial muscles, and high-capacitance dielectrics for plastic electronics and energy storage.

For example, consider ion gels, which are polymeric networks swollen with an ionic liquid (*2*). An ABA triblock copolymer (where the B block is soluble in the ionic liquid but the A blocks are not) can self-assemble at concentrations of a few percent by volume to provide a soft but highly elastic solid with a mesh size of 10 to 100 nm. The polymer network thus acts like a sponge, with the holes entirely filled with ionic liquid; because the holes are much larger than the ions, ion mobilities are comparable to those in the neat ionic liquid (see the first figure, center) (*3*). For a suitable A block–ionic liquid combination, the A blocks can dissolve at elevated temperature, providing a thermally reversible solid-liquid transition that enables liquid-state processing but solid-state use (see the first figure, right) (*4*).

Department of Chemistry and Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: lodge@umn.edu

Published by AAAS

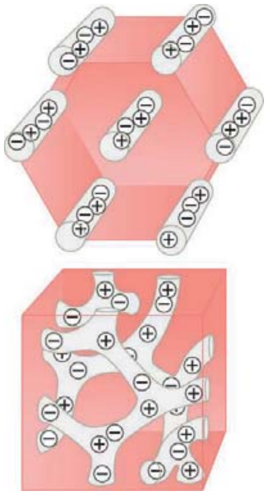
Such ion gels have shown superior performance as gate dielectrics in organic thin-film transistors (*5*). For many organic semiconductors, device performance is constrained by the number of charge carriers rather than their mobility; the high capacitance of the ion gel boosts the carrier density in the semiconductor channel. Further, the high ionic mobility enables switching speeds that are orders of magnitude higher than with conventional polymer electrolytes (*6*).

Similar ion gels could form the basis of electromechanical actuators (*7*, *8*); differential ion migration in response to an applied electric field leads to differential gel swelling and thus to bending. A possible route to accentuate this effect would be to polymerize the cations into the B blocks, thus immobilizing some or all of that charge (*9*, *10*); the much more mobile anions could then generate a highly asymmetric swelling.

The same molecular architecture also holds promise for gas separation. Ionic liquids strongly prefer to dissolve CO_2 and SO_2 over, for example, N_2 and CH_4. Because transport through an ionic liquid is so facile, it is possible to achieve combinations of selectivity and throughput comparable to those of the best materials currently available. However, a functional gas separation membrane must withstand a substantial pressure drop. The ionic liquid could be literally blown out of the simple ion gel. A polymerized ion gel should not suffer from this drawback, because the attraction between ions would far outweigh the external pressure. Direct polymerization of organic cations has recently been achieved (*9*, *10*). By incorporating an appropriate difunctional monomer, Bara *et al.* have prepared and evaluated cross-linked films for the separation of CO_2 from CH_4 or N_2 (*9*), with promising results.

Applications to other technologies such as fuel cell membranes and lithium battery separators often require much greater mechanical rigidity and high-temperature stability while retaining high ionic mobility along a given axis. Here, the ability of block polymers to self-assemble into well-defined nanostructures with long-range order holds the key (*11*). For example, macroscopic orientation of block polymer cylinders has been achieved by various strategies, including application of flow fields and electric fields and by preparation of suitably treated underlying substrates (see the second figure, top panel) . However, it remains difficult to achieve macroscopic orientation and perfection of the resulting membrane, which is important for some applications. Alternatively, a network structure such as the double gyroid is isotropic, obviating the need for orientation (see the second figure, bottom panel). Unfortunately, this structure can only be achieved under limited combinations of copolymer compositions, molar masses, and processing conditions. Use of multiblock polymers, such as ABC terpolymers, allows network materials to be prepared over much wider ranges of molecular variables and with greatly enhanced mechanical strength (*12*).

Recent progress in the development of controlled polymerization has enabled the synthesis of almost any desired architecture, with almost unlimited choice of monomers, such that tailored multiblock polymers can be readily produced. The outstanding challenges are to design block polymer–ionic liquid composite materials with desired combinations of mechanical integrity, controlled nanostructure, and ionic liquid properties. This class of materials is only beginning to be explored and many design rules are yet to be mapped out in detail, but rapid progress can be expected because related polymer-based technologies are well established.



Self-assembly. Block polymers can self-assemble into (**top**) hexagonally packed cylinders or (**bottom**) the double gyroid. The ionic liquid is confined to the white channels, whereas the red matrix consists of the insoluble block.

CREDIT: ADAPTED FROM TIMOTHY P. LODGE

References

1. T. Ueki, M. Watanabe, *Macromolecules* **41**, 3739 (2008).
2. M. A. B. H. Susan, T. Kaneko, A. Noda, M. Watanabe, *J. Am. Chem. Soc.* **127**, 4976 (2005).
3. Y. He, P. G. Boswell, P. Bühlmann, T. P. Lodge, *J. Phys. Chem. B* **111**, 4645 (2007).
4. Y. He, T. P. Lodge, *Macromolecules* **41**, 167 (2008).
5. J. Lee, M. J. Panzer, Y. He, T. P. Lodge, C. D. Frisbie, *J. Am. Chem. Soc*. **129**, 4532 (2007).
6. J. H. Cho *et al*., *Adv. Mater.* **20**, 686 (2008).
7. W. Lu *et al*., *Science* **297**, 983 (2002).
8. J. Ding *et al*., *Chem. Mater.* **15**, 2392 (2003).
9. J. E. Bara *et al*., *Ind. Eng. Chem*. **46**, 5397 (2007).
10. J. Tang *et al*., *Chem. Commun*. **2005**, 3325 (2005).
11. P. M. Simone, T. P. Lodge, *Macromolecules* **41**, 1753 (2008).
12. T. H. Epps III *et al*., *Macromolecules* **37**, 7085 (2004).

10.1126/science.1159652

OCEANS

Carbon Emissions and Acidification

Richard E. Zeebe[1*], James C. Zachos,[2] Ken Caldeira,[3] Toby Tyrrell[4]

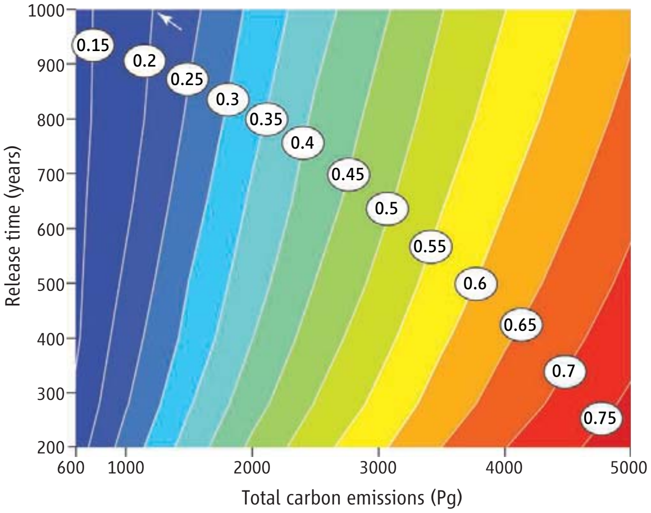
Avoiding environmental damage from ocean acidification requires reductions in carbon dioxide emissions regardless of climate change.

Much of the scientific and public focus on anthropogenic carbon dioxide (CO_2) emissions has been on climate impacts. Emission targets have been suggested based primarily on arguments for preventing climate from shifting significantly from its preindustrial state. However, recent studies underline a second major impact of carbon emissions: ocean acidification. Over the past 200 years, the oceans have taken up ~40% of the anthropogenic CO_2 emissions. This uptake slows the rise in atmospheric CO_2 considerably, thus alleviating climate change caused by anthropogenic greenhouse gas emissions. But it also alters ocean chemistry, with potentially serious consequences for marine life (*1*).

Oceanic uptake of anthropogenic CO_2 leads to a decrease in seawater pH and thus lowers the saturation state for carbonate minerals such as calcite and aragonite ($CaCO_3$). This process, termed ocean acidification, is expected to have detrimental consequences for a variety of marine organisms (*2*, *3*). For example, a decline in carbonate saturation state will affect the stability and likely production rates of $CaCO_3$ minerals, which are the building blocks of coral reefs and form the shells and skeletons of other marine calcifying species. Independent of climatic considerations, carbon emissions must be reduced to avoid these consequences.

[1] Department of Oceanography, University of Hawaii, Honolulu, HI 96822, USA. [2] Earth and Planetary Sciences Department, University of California at Santa Cruz, Santa Cruz, CA 95060, USA. [3] Department of Global Ecology, Carnegie Institution, Stanford, CA 94305, USA. [4] National Oceanography Centre, Southampton University, Southampton, SO14 3ZH, UK. *To whom correspondence should be addressed. E-mail: zeebe@soest.hawaii.edu

The range of tolerable pH changes (where "tolerable" means "without substantial detriment to organism fitness") is as yet unknown for many marine organisms. In laboratory and mesocosm studies, a decrease of 0.2 to 0.3 units in seawater pH inhibits or slows calcification in many marine organisms, including corals (*4*), foraminifera (*5*), and some calcareous plankton (*6*, *7*). A drop of 0.3 pH units corresponds to a doubling of the hydrogen ion concentration $[H^+]$, because pH is expressed on a logarithmic scale as $pH = -\log([H^+])$.



Surface ocean pH decline. The white contour lines illustrate the expected maximum pH decrease of average surface ocean waters in the future (in pH units) as a function of total anthropogenic CO_2 emissions (in petagrams of carbon, 1 Pg = 10^{15} g) and release time (in years, see supporting online material). For example, if humans release a total of 1200 Pg C over 1000 years, surface ocean pH will drop by about 0.2 units (arrow).

Compared with preindustrial levels, average surface ocean pH has already decreased by ~0.1 units (*2*). If future increases in seawater acidity affect calcification in coral reefs, such that erosion outweighs accretion, then the reefs could lose structural stability, with further negative implications for reef communities and for shore protection (*3*). Reduced calcification in shellfish such as oysters and mussels would impact worldwide commercial aquaculture production (*8*). Effects of ocean acidification on noncalcifying organisms such as viruses and bacteria are largely unknown, as are potential consequences for marine food webs.

Thus, although the response of different organisms is expected to be inhomogeneous (*9*), current evidence suggests that large and rapid changes in ocean pH will have adverse effects on a number of marine organisms. Yet, environmental standards for tolerable pH changes have not been updated in decades. For example, the seawater quality criteria of the U.S. Environmental Protection Agency date back to 1976 and state that for marine aquatic life, pH should not be changed by more than 0.2 units outside of the normally occurring range (*10*). These standards must be reevaluated based on the latest research on pH effects on marine organisms. Once new ranges of tolerable pH are adopted, CO_2 emission targets must be established to meet those requirements in terms of future seawater chemistry changes.

The key variables that will determine the extent of future seawater chemistry changes—and therefore the impact on marine life—is the magnitude and time scale of the anthropogenic carbon release. For specific CO_2 emission scenarios, changes in atmospheric CO_2, surface ocean pH, and carbonate mineral saturation state have been calculated with different carbon cycle models (*1*, *11*–*13*). Fortunately, and in contrast to climate model predictions, such future ocean chemistry projections are largely model-independent on a time scale of a few centuries, mainly because the chemistry of CO_2 in seawater is well known (*14*) and changes in surface ocean carbonate chemistry closely track changes in atmospheric CO_2. Predicted changes in surface ocean pH for given total emissions and release time of anthropogenic carbon are thus similar among different types of models over the next few centuries [see supporting online material and (*1*, *11*–*13*)].

Projected changes in ocean carbonate chemistry should serve as a guideline for policy protocols that identify CO_2 emission targets to reduce the effects of human-made ocean acidification. For example, to avoid a surface ocean pH decline by more than 0.2 units (*10*), total emission targets would have to range from ~700 Pg C over 200 years to ~1200 Pg C over 1000 years (see the figure). Such scenarios would be difficult to achieve, however, because they require immediate reductions in global emissions. If emissions can be reduced after the year 2050 and capped at 1500 Pg C, surface ocean pH would decline by ~0.35 units relative to preindustrial levels. The aragonite saturation state in the warm surface ocean would drop from ~3.5 to ~2.1 under this scenario (see supporting online material). Substantial reductions in coral calcification have been reported over this range (*2*–*4*).

Predictions of changes in calcification rates and other physiological responses in marine organisms to ocean acidification are hampered by a lack of adequate experimental data and field observations. Most studies indicate reduced calcification rates at high CO_2 (*2*–*8*, *11*), but some find little change or increasing cellular calcification rates (*9*, *15*). The key is to understand the response of the functional groups that drive marine biogeochemical cycles. For instance, very few studies have examined foraminifera (*5*), which are major contributors to the production and deposition of calcium carbonate in the ocean. Also, long-term experiments are necessary with different calcifying and non-calcifying groups to test their ability to adapt to a high–CO_2 world.

To monitor and quantify future changes in ocean chemistry and biogeochemical fluxes, intensified global-ocean carbon dioxide surveys in combination with carbon-cycle modeling will be necessary. Awareness must be raised among the public and policy-makers of the effects of ocean acidification and the steps required to control it. Ocean chemistry changes, and not only climate effects, should be taken into consideration when determining CO_2 emission targets; such consideration is likely to weigh in favor of lower emission targets.

References

1. K. Caldeira, M. E. Wickett, *Nature* **425**, 365 (2003).
2. J. Raven *et al.*, *Ocean Acidification Due to Increasing Atmospheric Carbon Dioxide* (The Royal Society, Policy Document, London, UK, 2005).
3. O. Hoegh-Guldberg *et al.*, *Science* **318**, 1737 (2007).
4. J. A. Kleypas *et al.*, *Science* **284**, 118 (1999).
5. J. Bijma, H. J. Spero, D. W. Lea, *Use of Proxies in Paleoceanography: Examples from the South Atlantic*, G. Fischer, G. Wefer, Eds. (Springer, Berlin, 1999), pp. 489–512.
6. D. A. Wolf-Gladrow, U. Riebesell, S. Burkhardt, J. Bijma, *Tellus* B **51**, 461 (1999).
7. U. Riebesell *et al.*, *Nature* **407**, 364 (2000).
8. F. Gazeau, C. Quiblier, J. M. Jansen, J.-P. Gatusso, *Geophys. Res. Lett.* **34**, L07603 (2007).
9. G. Langer *et al.*, *Geochem. Geophys. Geosyst.* **7**, Q09006 (2006).
10. U.S. Environmental Protection Agency, *Quality Criteria for Water* (Washington, DC, 1976).
11. J. C. Orr *et al.*, *Nature* **437**, 681 (2005).
12. A. Montenegro, V. Brovkin, M. Eby, D. Archer, A. J. Weaver, *Geophys. Res. Lett.* **34**, L19707 (2007).
13. J. C. Zachos, G. R. Dickens, R. E. Zeebe, *Nature* **451**, 279 (2008).
14. R. E. Zeebe, D. A. Wolf-Gladrow, *CO_2 in Seawater: Equilibrium, Kinetics, Isotopes* (Elsevier Oceanography Series, Amsterdam, 2001).
15. M. D. Iglesias-Rodriguez *et al.*, *Science* **320**, 336 (2008).
16. This work was supported by NSF grant NSFOCE07-51959 to R.E.Z.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/51/DC1
Figs. S1 and S2
References

10.1126/science.1159124

Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations

Thomas Klausberger[1,2]* and Peter Somogyi[1]*

In the cerebral cortex, diverse types of neurons form intricate circuits and cooperate in time for the processing and storage of information. Recent advances reveal a spatiotemporal division of labor in cortical circuits, as exemplified in the CA1 hippocampal area. In particular, distinct GABAergic (γ-aminobutyric acid–releasing) cell types subdivide the surface of pyramidal cells and act in discrete time windows, either on the same or on different subcellular compartments. They also interact with glutamatergic pyramidal cell inputs in a domain-specific manner and support synaptic temporal dynamics, network oscillations, selection of cell assemblies, and the implementation of brain states. The spatiotemporal specializations in cortical circuits reveal that cellular diversity and temporal dynamics coemerged during evolution, providing a basis for cognitive behavior.

The cerebral cortex of mammals has a large diversity of cells operating in intricate circuits. This cellular diversity endows the cerebral cortex with the capacity to perform complex biological processes such as the subjective representation and interpretation of the world, encoding and retrieval of emotionally colored memories, understanding and empathizing with other individuals, and scientifically investigating the universe (including the mind). We argue here that time is the key metric to all cortical operations. Temporal demands drive selection for computational sophistication and also drive the evolution of neuronal diversity. In turn, cellular diversity serves the temporal organization of cortical functions in the coordination of the activity of different subcellular domains of a single neuron as well as neuronal populations.

The exploration of different cells started in the late 19th century, enduringly represented by Ramon y Cajal's insights into connectivity through Golgi impregnation, which revealed the processes of single neurons (*1*). To the embarrassment of thousands of neuroscientists trying to explain cortical events through specific circuits today, we still lack the basic knowledge of how many types of neuron exist and how cells are interconnected. The CA1 area of the hippocampus constitutes one of the simplest and most examined cortical areas where recent progress has been made in explaining neuronal diversity and the temporal activity of distinct cells. Here, excitatory pyramidal cells encode representations of spatial (*2*) and other episodic memories (*3*) and provide glutamatergic output to other cortical as well as subcortical areas. Although few differences have been noted in CA1 pyramidal cells, they (Fig. 1, neuron types 22 to 24) represent at least three distinct types (fig. S1) targeting more than 10 extrahippocampal brain areas (*4*). The variety of areas targeted by each individual cell remains to be established.

The relatively uniform pyramidal cells are supported by a rich diversity of GABAergic interneurons that provide general inhibition and also temporally regulate pyramidal cell activity. Interneurons are recognized on the basis of firing patterns, molecular expression profiles, and their innervations of distinct subcellular domains of pyramidal cells (Fig. 1). The GABAergic interneuron types are not unique to the CA1 area; similar neurons are present in most other areas of the hippocampus and the isocortex (*5*). Furthermore, these GABAergic interneurons can be found in mouse, rat, cat, monkey, and the human cortex. Why has such a highly structured neuronal machinery evolved and been preserved throughout evolution? Why do GABAergic interneurons, rather than glutamatergic principal cells, show the largest cellular diversity? We explore three answers to these questions and discuss how the dynamic timing of synaptic action between different types of interneuron and pyramidal cells supports distinct brain states and cognitive processing.

The Soma, Axon-Initial Segment, and Distinct Dendritic Domains of Pyramidal Cells Receive GABAergic Innervations Differentiated in Time

A CA1 pyramidal cell receives about 30,000 synaptic inputs and emits several types of dendrite to provide a framework for their integration. The cell body integrates inputs from the dendrites and receives only GABAergic synapses, as does the axon-initial segment, which contributes to action potential generation. The small, oblique dendrites emerging from one or two large apical dendrites and the basal dendrites receive glutamatergic input mainly from the hippocampal CA3 area, local axon collaterals, and the amygdala. The apical dendritic tuft is innervated mainly by glutamatergic inputs from the entorhinal cortex and the thalamus. All dendrites also receive local GABAergic inputs from interneurons. Such a compartmentalized structure of pyramidal cells allows spatially segregated activities at the same time. Interestingly, different types of parvalbumin (PV)–expressing, GABAergic interneuron also innervate distinct subcellular domains: Axo-axonic cells (Fig. 1, type 1) innervate exclusively the axon-initial segment of pyramidal cells; basket cells (Fig. 1, type 2) innervate the cell bodies and proximal dendrites; bistratified cells (Fig. 1, type 5) innervate the basal and oblique dendrites coaligned with the CA3 glutamatergic input; and oriens–lacunosum moleculare (O-LM) interneurons (Fig. 1, type 7) target the apical dendritic tuft aligned with the entorhinal cortical input.

Indications that interneurons might contribute differentially to the temporal coordination of pyramidal cells came from in vitro recordings in brain slices of rats. Perisomatic innervating interneurons modulate the probability of sodium spikes, and some dendritic GABAergic innervation interferes with Ca^{2+}-dependent spike generation (*6*). Furthermore, differences in the short-term plasticity of glutamatergic synapses onto distinct interneurons (*7–10*) may lead to a temporally distinct and spatially distributed recurrent inhibition in perisomatic or dendritic domains of pyramidal cells (*11*). In the somatosensory cortex, the firing frequency of perisomatic- and dendrite-targeting interneurons may differentially entrain the output of postsynaptic pyramidal cells (*12*). The dissection of cellular properties in vitro has provided stimulating possibilities of how distinct types of interneuron might act. It remains a challenge to explain how these concepts relate to the information flow when the neurons are embedded in ongoing network activity.

A temporally distinct contribution of interneurons in the intact rat brain was indicated by diverse firing patterns of putative and unidentified interneurons during network oscillations (*13*). Network oscillations in the cerebral cortex indicate highly coordinated neuronal activity (*14*) over large areas. For example, theta oscillations (4 to 10 Hz) highlight the online state of the hippocampus and related structures. Theta waves together with gamma oscillations (30 to 80 Hz) occur during spatial navigation, memory tasks, and rapid-eye-movement sleep. In contrast, sharp wave-associated ripples (100 to 200 Hz) occur during resting, consummatory behavior, and slow-wave sleep, supporting offline replay and consolidation of previous experiences (*15*, *16*). The spike timing of putative interneurons can be referenced to the network events (*13*). The recording of identified interneurons in anesthetized rats demonstrated that interneurons belonging to distinct classes defined by their axonal target domain

[1]MRC Anatomical Neuropharmacology Unit, Oxford University, Oxford OX1 3TH, UK. [2]Center for Brain Research, Medical University of Vienna, A-1090 Vienna, Austria.

*To whom correspondence should be addressed. E-mail: thomas.klausberger@pharm.ox.ac.uk (T.K.); peter.somogyi@pharm.ox.ac.uk (P.S.)

on the pyramidal cell do indeed fire action potentials at distinct times (Fig. 2).

During ripple oscillations, basket (*17*) and bistratified cells (*18*) strongly increase their firing rate and discharge in a manner phase-coupled to the oscillatory cycles. In contrast, axo-axonic cells fire sometimes before the ripple episode but are silenced during and after it, and O-LM cell firing is suppressed during ripples (*19*). Because these different interneurons innervate distinct domains of pyramidal cells, they imprint a spatiotemporal GABAergic conductance matrix onto the pyramidal cells. This GABAergic fingerprint changes its pattern during different brain states. During theta oscillations, O-LM cells (*19*) become very active and, in cooperation with bistratified cells (*18*), modulate the dendrites of pyramidal cells one-quarter of a theta cycle after PV-expressing basket cells (*20*) discharge; PV-expressing basket cells in turn fire later than axo-axonic cells (*19*). Also, during gamma oscillations, distinct types of interneuron contribute differentially to the temporal modulation of pyramidal cell subcellular domains (Fig. 2) (*21*–*23*). Firing of basket (*24*) and axo-axonic cells (*23*) is moderately coupled to the ascending phase of extracellular gamma oscillations in the pyramidal cell layer. In contrast, spike timing of bistratified cells is most tightly correlated to field gamma, whereas O-LM cells do not contribute (*23*) to the synchronization of pyramidal cells to network gamma oscillations in the CA1 area.

In summary, the different classes of interneurons that have been tested fire action potentials, and presumably release GABA, at different time points to distinct subcellular domains of pyramidal cells. Therefore, GABA cannot be provided by the axon of a single type of neuron; instead, independently firing cell classes are required to support the distributed computations of pyramidal cells.

The Same Domain of Pyramidal Cells Receives Differentially Timed GABAergic Input from Distinct Sources

In addition to PV-expressing cells, cholecystokinin (CCK)–expressing GABAergic interneurons also innervate pyramidal cells (Fig. 1) at the soma and proximal dendrites (types 3 and 4), at the apical dendrites (type 9), at dendrites receiving glutamatergic CA3 input (type 8), and at the apical tuft (type 10). These CCK-expressing cells receive specific inputs from modulatory brainstem nuclei (*25*) and fire different spike trains in vitro (*26*); their asynchronous GABA release causes longer-lasting inhibition in pyramidal cells (*27*); and their inhibitory effect is attenuated by postsynaptic pyramidal cells via cannabinoid receptors (*28*). Electrical stimulation of presynaptic fibers in vitro indicated that CCK-expressing cells may be particularly suited for integrating excitation from multiple afferents (*29*).

In vivo recordings of identified CCK-expressing cells in anesthetized rats (*30*) showed that CCK- and PV-expressing interneurons fire at distinct times (Fig. 2). During theta oscillations, CCK-expressing cells fire at a phase when CA1 pyramidal cells start firing as the rat enters a spatial location, the place field of the cell. During gamma oscillations, CCK-expressing cells fire just before CA1 pyramidal cells (*23*). Because active pyramidal cells can selectively reduce the inhibition from CCK-expressing cells via retrograde cannabinoid receptor activation, the unique spike timing and molecular design of these GABAergic cells are well suited to increase the contrast in

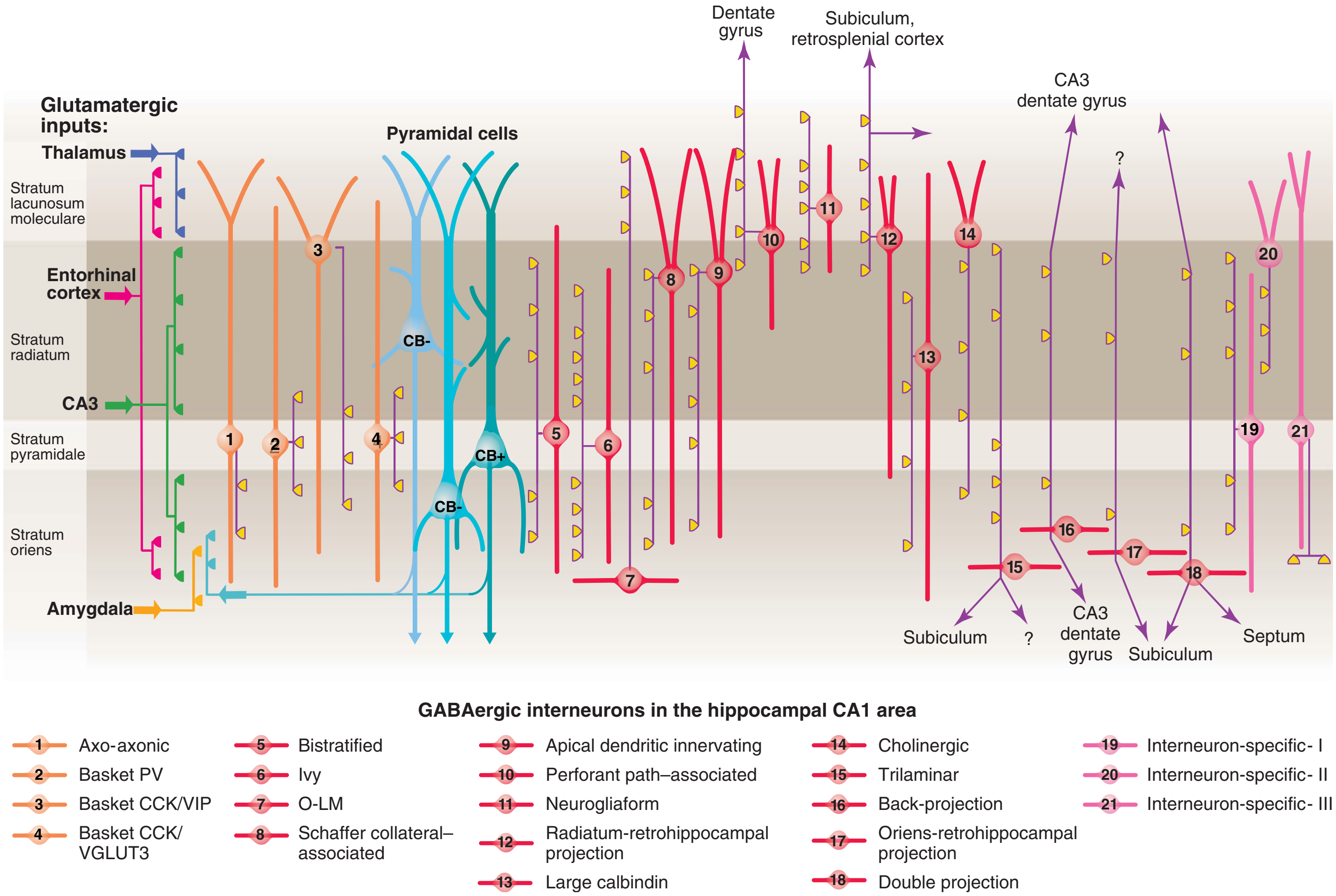


Fig. 1. Three types of pyramidal cell are accompanied by at least 21 classes of interneuron in the hippocampal CA1 area. The main termination of five glutamatergic inputs are indicated on the left. The somata and dendrites of interneurons innervating pyramidal cells (blue) are orange, and those innervating mainly other interneurons are pink. Axons are purple; the main synaptic terminations are yellow. Note the association of the output synapses of different interneuron types with the perisomatic region (left) and either the Schaffer collateral/commissural or the entorhinal pathway termination zones (right), respectively. VIP, vasoactive intestinal polypeptide; VGLUT, vesicular glutamate transporter; O-LM, oriens lacunosum moleculare.

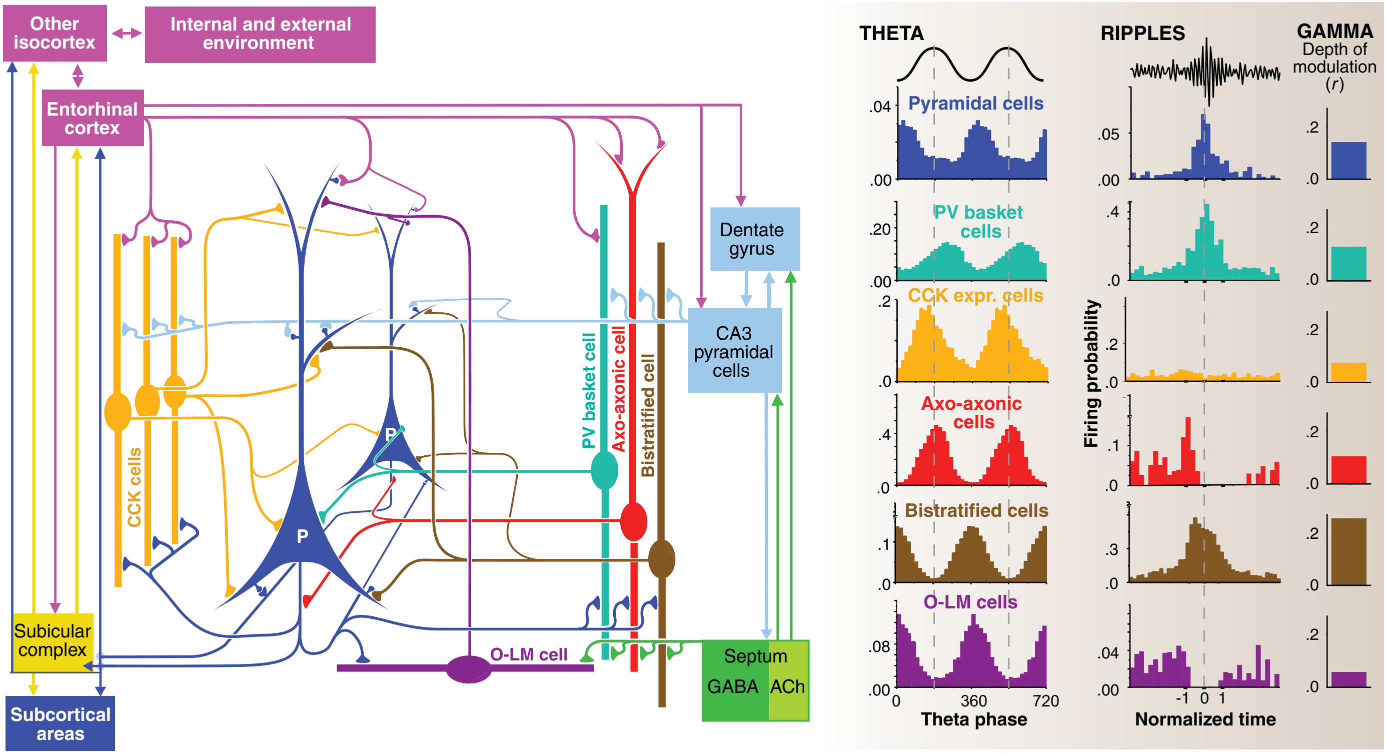


Fig. 2. Spatiotemporal interaction between pyramidal cells and several classes of interneuron during network oscillations, shown as a schematic summary of the main synaptic connections of pyramidal cells (P), PV-expressing basket, axo-axonic, bistratified, O-LM, and three classes of CCK-expressing interneurons. The firing probability histograms show that interneurons innervating different domains of pyramidal cells fire with distinct temporal patterns during theta and ripple oscillations, and their spike timing is coupled to field gamma oscillations to differing degrees. The same somatic and dendritic domains receive differentially timed input from several types of GABAergic interneuron (*18*, *19*, *23*, *30*). ACh, acetylcholine.

the firing of strongly active (disinhibited via CB1 receptors) and weakly active or inactive (still inhibited by CCK interneurons) pyramidal cells, supporting the implementation of sparse coding in cell assemblies. The sum of PV- and CCK-expressing basket cell activity, together with axo-axonic cell firing, is maximal when pyramidal cell firing is minimal during theta oscillations. The different spike timing of CCK- and PV-expressing interneurons is likely to be generated by synaptic inputs from distinct sources, thus demonstrating the cooperation of temporal and spatial organization.

In addition, the dendrites of pyramidal cells are also innervated by GABAergic neurogliaform cells, which provide slow GABA_A receptor–mediated (*31*, *32*) and also GABA_B receptor–mediated inhibition (*33*, *34*). Neurogliaform cells (type 11) innervate the apical dendritic tuft of CA1 pyramidal cells co-aligned with the entorhinal input, whereas a related cell type, the Ivy cell (type 6), innervates more proximal pyramidal cell dendrites aligned with the CA3 input (Fig. 1). The spatially complementary axonal termination of Ivy and neurogliaform cells is mirrored by distinct spike timing in vivo (*35*, *36*). Ivy cells expressing nitric oxide synthase and neuropeptide Y, but neither PV nor CCK, represent the most numerous class of interneuron described so far. They evoke slow GABAergic inhibition in pyramidal cells, and through neuropeptide Y signaling they are likely to modulate glutamate release from terminals of CA3 pyramidal cells, which, in contrast to perforant path terminals, express a high level of Y2 receptor (*37*). Ivy cells, together with neurogliaform cells, are a major source of nitric oxide, probably released by their extraordinarily dense axons. They modulate pre- and postsynaptic excitability at slower time scales and more diffusely than do other interneurons providing homeostasis to the network.

How the different firing patterns of distinct GABAergic neurons are generated remains largely unknown. For example, since the discovery of axo-axonic cells in 1983 (*38*), only one glutamatergic input from CA1 pyramidal cells has been published (*39*); all other excitatory and inhibitory inputs remain inferential predictions. Potential candidates for governing the activation of interneurons include differential glutamatergic and subcortical innervation (*40*, *41*), selective GABAergic and electrical coupling between interneurons, cell type–specific modulatory regulation (*42*), cell type–specific expression of distinct receptors and channels (*43*–*46*), or differential input from interneurons (Fig. 1, types 19 to 21), which apparently innervate exclusively other interneurons (*47*, *48*). Little is known about the activity of the latter cell types in vivo. Interestingly, the only subcellular pyramidal cell domain that receives GABAergic input from a single source is the axon-initial segment, which highlights the unique place of axo-axonic cells in the cortex of mammals.

The Coordination of Network States Across Cortical Areas Is Supported by GABAergic Projection Neurons

Many distributed areas of the cerebral cortex participate in each cognitive process. Coordination is supported by shared subcortical pathways and by inter-areal pyramidal cell projections terminating on both pyramidal cells and local GABAergic interneurons. In addition, GABAergic corticocortical connections are also present [e.g., (*49*)], including those in the temporal lobe (*50*). Some neurons (Fig. 1, type 16) project to neighboring hippocampal subfields (*51*) and/or to the medial septum (type 18) (*52*), a key structure regulating network states. Recording and labeling GABAergic neurons in vivo revealed a variety of GABAergic projection neurons (*50*). Hippocamposeptal neurons (type 18) also send thick, myelinated axons to the subiculum and other retrohippocampal areas; other GABAergic cells (types 15 and 17) project only to retrohippocampal areas, parallel with glutamatergic CA1 pyramidal cells. Because these projection cells fire rhythmically during sharp wave-associated ripple and gamma oscillations, they contribute to temporal organization across the septohippocampal-subicular circuit. In addition, other GABAergic projection neurons (Fig. 1, type 12) emit long-

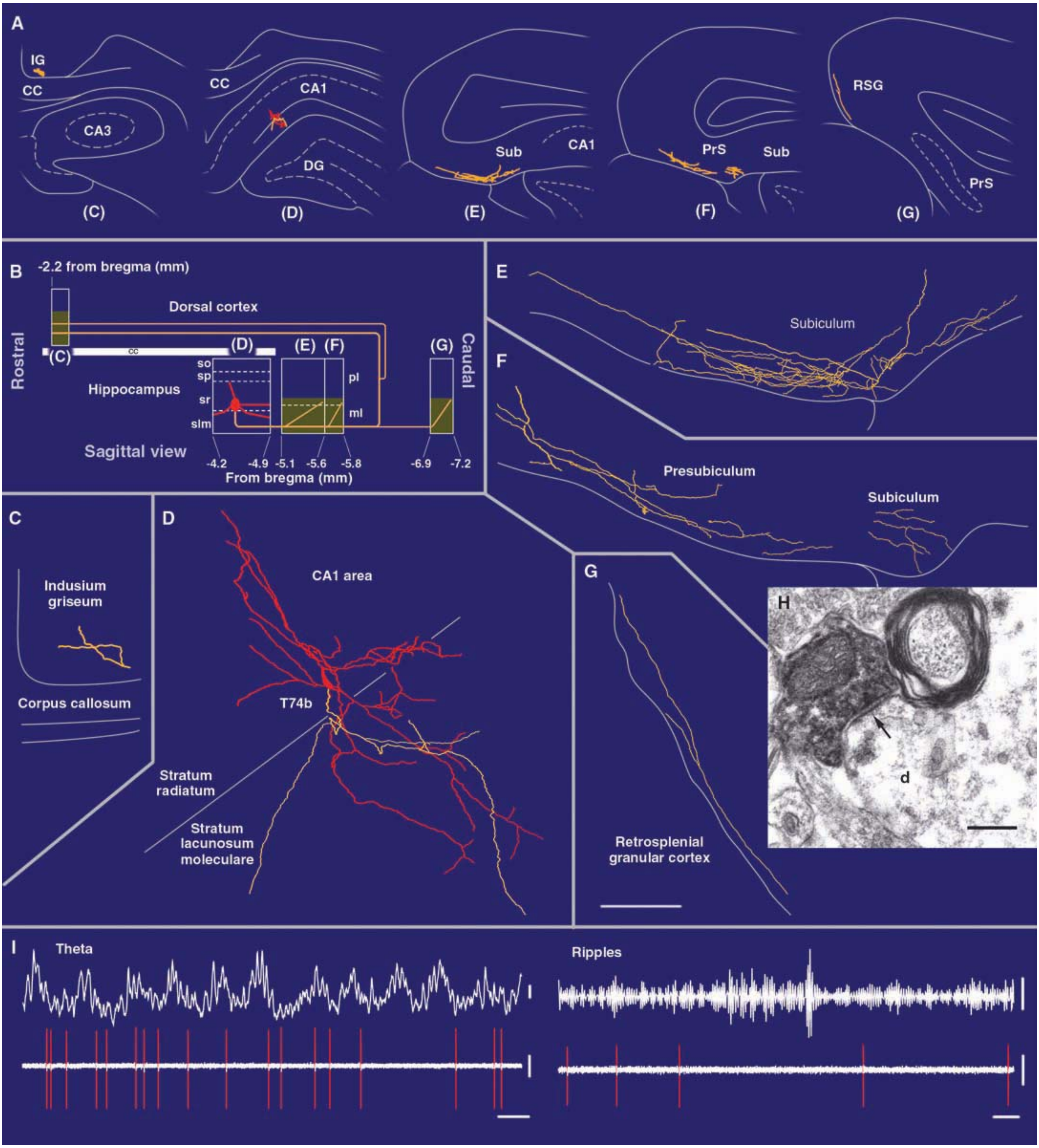


Fig. 3. In vivo spike timing of a GABAergic CA1 neuron projecting to the subiculum (Sub), presubiculum (PrS), retrosplenial cortex (RSG), and indusium griseum (IG). (**A**) The soma, dendrites (red), and axons (yellow) in coronal plains as indicated in (B). CC, corpus callosum. (**B**) Representation in the sagittal plane, showing the rostrocaudal extent of the cell. The soma is located at the border of the stratum radiatum and lacunosum moleculare. The axon, traced over 5 mm, runs toward caudal regions through the subiculum and presubiculum, then bifurcates into further caudal and rostral branches. Shaded areas represent boutons in the reconstructed sections. (**C** to **G**) Soma and dendrites are complete; the axon is shown from selected sections [blocks in (B)]; note few local collaterals within the hippocampus. The axon innervates the molecular layer in the subiculum and the retrosplenial granular cortex. (**H**) Electron micrograph of a neurobiotin-filled bouton making a type II synapse (arrow) with a dendritic shaft in the subiculum. (**I**) In vivo firing patterns show that the cell fires at the descending phase of extracellular theta oscillations (filter, direct current to 220 Hz) recorded from a second electrode in the pyramidal layer. During ripple episodes (right upper, 90 to 140 Hz band pass), there is no increase in firing. Scale bars, 100 μm [(C) to (G)], 0.2 μm (H). Calibrations in (I): theta, 0.2 mV; ripples, 0.05 mV, 0.1 s; spikes, 0.5 mV.

CREDIT: REPRODUCED FROM (50) WITH PERMISSION

range myelinated axons that arborize in the molecular layers of subiculum, presubiculum, and retrosplenial cortex (*50*). Their rhythmic firing during theta oscillations indicates a contribution to the temporal organization of this brain state across the targeted areas (Fig. 3).

In summary, information between cortical areas is transmitted via axonal projections of glutamatergic pyramidal cells, but they alone may not produce the required high degree of temporal precision between brain regions. Together with common subcortical state-modulating inputs, the cortical long-range GABAergic projections could prime and reset activity in specific neurons of the target areas before the information arrives via glutamatergic fibers from pyramidal cells. The neuronal diversity of GABAergic projection neurons differing in target area and temporal activity increases computational powers between related cortical areas.

Dynamic Cooperation of Pyramidal Cells and Specific GABAergic Interneurons in Cell Assemblies

The diversity of subcellular domain-specific GABAergic interneurons, combinatorial input to the very same subcellular domain, and differentiated projections to long-range target areas provide a basis for dynamic, rather than clock-like, regulation of pyramidal cell networks. Without this cooperative temporal framework, the glutamatergic connections would lose meaning. Indeed, the stimulation of a single putative interneuron in the barrel cortex can affect behavioral responses (*53*), and hippocampal interneurons actively participate in recognition memory (*54*). Although many interneurons fire at high rates during theta oscillations, some cells also show increased firing when the animal is in a particular location (*55*, *56*), a hallmark of pyramidal place cells (*2*). Like pyramidal cells, some interneurons also fire at progressively earlier phases of the theta cycles when the rat passes through fields of increased firing (*57*–*59*). Together with the observation that CCK-expressing and axo-axonic cells fire only during and before some of the ripple episodes (*19*, *30*), this indicates that some GABAergic interneurons contribute to the dynamic selection and control of cell assemblies. Such an interneuron contribution is not reflected in linear flowcharts of synfire chains, in which interneurons simply inhibit pyramidal cells that are not part of that assembly or delay the firing of pyramidal cells, which fire later in the chain.

As GABAergic interneurons innervate thousands of nearby pyramidal cells, a hard-wiring that includes GABAergic connections is unlikely, given the large number of representations involving the same pyramidal cell (*60*). The effect of domain-specific GABAergic inputs is more likely dependent on the state of the receiving pyramidal cells or the domains on a single pyramidal cell. For example, axo-axonic cells might depolarize the axon-initial segment depending on the membrane potential, momentary internal chloride concentration, and recent history of ion channels in the postsynaptic membrane (*61*), whereas cross-correlation of firing patterns in vivo (Fig. 2) suggests that axo-axonic cells, on average, inhibit CA1 pyramidal cells. The GABAergic input to a dendrite might shunt other inputs, de-inactivate voltage-gated cation channels through hyperpolarization, reset the phase of intrinsic dendritic oscillations, or gate the incoming excitation in a winner-take-all manner, according to the recent history of that dendrite. Such state-dependent effects, together with the powerful combinatorial GABAergic inputs from independently regulated cell classes, enable the nonlinear emergence of cell assemblies.

Conclusion

Recording the spike timing of identified neurons has revealed that the large diversity of cortical neurons is accompanied by an equally sophisticated temporal differentiation of their activity. Indeed, the existence of one can only be explained and understood by the other. The in vivo firing patterns also indicate that a classification of GABAergic interneurons based on axonal target specificity and molecular expression profile correctly groups cells according to their temporal contribution to network activity. It remains to be seen whether the differential expression of calbindin by CA1 pyramidal cells is also accompanied by different temporal (*62*) and axonal target specialization and by a distinct contribution to cognitive operations. Defining all neuronal populations, their molecular expression profile (*63*), synaptic connections, and temporal activity, together with changing the activity of selected cell types or connections (*64*), will explain cortical circuits and how defects in timing cause cortical pathologies (*65*–*68*).

References and Notes

1. S. Ramon y Cajal, *Histologie du Systeme Nerveux de l'Homme et des Vertebres* (Maloine, Paris, 1911), chap. II.
2. J. O'Keefe, *Exp. Neurol.* **51**, 78 (1976).
3. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, *Nature* **435**, 1102 (2005).
4. L. A. Cenquizca, L. W. Swanson, *J. Comp. Neurol.* **497**, 101 (2006).
5. J. S. Lund, D. A. Lewis, *J. Comp. Neurol.* **328**, 282 (1993).
6. R. Miles, K. Toth, A. I. Gulyas, N. Hajos, T. F. Freund, *Neuron* **16**, 815 (1996).
7. A. B. Ali, J. Deuchars, H. Pawelzik, A. M. Thomson, *J. Physiol.* **507**, 201 (1998).
8. A. B. Ali, A. M. Thomson, *J. Physiol.* **507**, 185 (1998).
9. A. A. Biro, N. B. Holderith, Z. Nusser, *J. Neurosci.* **25**, 223 (2005).
10. G. Silberberg, C. Wu, H. Markram, *J. Physiol.* **556**, 19 (2004).
11. F. Pouille, M. Scanziani, *Nature* **429**, 717 (2004).
12. G. Tamas, J. Szabadics, A. Lorincz, P. Somogyi, *Eur. J. Neurosci.* **20**, 2681 (2004).
13. J. Csicsvari, H. Hirase, A. Czurko, A. Mamiya, G. Buzsaki, *J. Neurosci.* **19**, 274 (1999).
14. I. Soltesz, M. Deschenes, *J. Neurophysiol.* **70**, 97 (1993).
15. D. J. Foster, M. A. Wilson, *Nature* **440**, 680 (2006).
16. K. Diba, G. Buzsaki, *Nat. Neurosci.* **10**, 1241 (2007).
17. A. Ylinen *et al.*, *J. Neurosci.* **15**, 30 (1995).
18. T. Klausberger *et al.*, *Nat. Neurosci.* **7**, 41 (2004).
19. T. Klausberger *et al.*, *Nature* **421**, 844 (2003).
20. A. Ylinen *et al.*, *Hippocampus* **5**, 78 (1995).
21. N. Hajos *et al.*, *J. Neurosci.* **24**, 9127 (2004).
22. T. Gloveli *et al.*, *J. Physiol.* **562**, 131 (2005).
23. J. J. Tukker, P. Fuentealba, K. Hartwich, P. Somogyi, T. Klausberger, *J. Neurosci.* **27**, 8184 (2007).
24. M. Penttonen, A. Kamondi, L. Acsady, G. Buzsaki, *Eur. J. Neurosci.* **10**, 718 (1998).
25. E. C. Papp, N. Hajos, L. Acsady, T. F. Freund, *Neuroscience* **90**, 369 (1999).
26. H. Pawelzik, D. I. Hughes, A. M. Thomson, *J. Comp. Neurol.* **443**, 346 (2002).
27. S. Hefft, P. Jonas, *Nat. Neurosci.* **8**, 1319 (2005).
28. I. Katona *et al.*, *J. Neurosci.* **19**, 4544 (1999).
29. L. L. Glickfeld, M. Scanziani, *Nat. Neurosci.* **9**, 807 (2006).
30. T. Klausberger *et al.*, *J. Neurosci.* **25**, 9782 (2005).
31. J. B. Hardie, R. A. Pearce, *J. Neurosci.* **26**, 8559 (2006).
32. J. Szabadics, G. Tamas, I. Soltesz, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14831 (2007).
33. G. Tamás, A. Lörincz, A. Simon, S. Szabadics, *Science* **299**, 1902 (2003).
34. C. J. Price *et al.*, *J. Neurosci.* **25**, 6775 (2005).
35. P. Fuentealba *et al.*, *Neuron* **57**, 917 (2008).
36. P. Fuentealba, T. Klausberger, P. Somogyi, unpublished data.
37. D. Stanic *et al.*, *J. Comp. Neurol.* **499**, 357 (2006).
38. P. Somogyi, M. G. Nunzi, A. Gorio, A. D. Smith, *Brain Res.* **259**, 137 (1983).
39. P. Ganter, P. Szücs, O. Paulsen, P. Somogyi, *Hippocampus* **14**, 232 (2004).
40. B. Kocsis, V. Varga, L. Dahan, A. Sik, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1059 (2006).
41. Z. Borhegyi, V. Varga, N. Szilagyi, D. Fabo, T. F. Freund, *J. Neurosci.* **24**, 8470 (2004).
42. C. Foldy, S. Y. Lee, J. Szabadics, A. Neu, I. Soltesz, *Nat. Neurosci.* **10**, 1128 (2007).
43. J. J. Lawrence, J. M. Statland, Z. M. Grinspan, C. J. McBain, *J. Physiol.* **570**, 595 (2006).
44. R. Cossart *et al.*, *Hippocampus* **16**, 408 (2006).
45. N. Hajos, I. Mody, *J. Neurosci.* **17**, 8427 (1997).
46. C. C. Lien, P. Jonas, *J. Neurosci.* **23**, 2058 (2003).
47. L. Acsady, T. J. Gorcs, T. F. Freund, *Neuroscience* **73**, 317 (1996).
48. A. I. Gulyas, N. Hajos, T. F. Freund, *J. Neurosci.* **16**, 3397 (1996).
49. R. Tomioka *et al.*, *Eur. J. Neurosci.* **21**, 1587 (2005).
50. S. Jinno *et al.*, *J. Neurosci.* **27**, 8790 (2007).
51. A. Sik, A. Ylinen, M. Penttonen, G. Buzsaki, *Science* **265**, 1722 (1994).
52. K. Toth, Z. Borhegyi, T. F. Freund, *J. Neurosci.* **13**, 3712 (1993).
53. A. R. Houweling, M. Brecht, *Nature* **451**, 65 (2008).
54. S. P. Wiebe, U. V. Staubli, *J. Neurosci.* **21**, 3955 (2001).
55. W. B. Wilent, D. A. Nitz, *J. Neurophysiol.* **97**, 4152 (2007).
56. L. Marshall *et al.*, *J. Neurosci.* **22**, RC197 (2002).
57. A. P. Maurer, S. L. Cowen, S. N. Burke, C. A. Barnes, B. L. McNaughton, *J. Neurosci.* **26**, 13485 (2006).
58. V. Ego-Stengel, M. A. Wilson, *Hippocampus* **17**, 161 (2007).
59. C. Geisler, D. Robbe, M. Zugaro, A. Sirota, G.Buzsáki, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8149 (2007).
60. S. Leutgeb *et al.*, *Science* **309**, 619 (2005).
61. J. Szabadics *et al.*, *Science* **311**, 233 (2006).
62. T. Jarsky, R. Mady, B. Kennedy, N. Spruston, *J. Comp. Neurol.* **506**, 535 (2008).
63. K. Sugino *et al.*, *Nat. Neurosci.* **9**, 99 (2006).
64. E. C. Fuchs *et al.*, *Neuron* **53**, 591 (2007).
65. G. T. Konopaske, R. A. Sweet, Q. Wu, A. Sampson, D. A. Lewis, *Neuroscience* **138**, 189 (2006).
66. I. Khalilov, M. LevanQuyen, H. Gozlan, Y. BenAri, *Neuron* **48**, 787 (2005).
67. I. Cohen, V. Navarro, S. Clemenceau, M. Baulac, R. Miles, *Science* **298**, 1418 (2002).
68. R. D. Traub *et al.*, *J. Neurophysiol.* **94**, 1225 (2005).
69. We thank the colleagues whose work laid the foundations for this review but could not be cited because of space restrictions. We also thank M. Capogna, N. Hajos, I. Katona and Z. Nusser for commenting on an earlier manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/53/DC1
Fig. S1
References

10.1126/science.1149381

MESSENGER

CONTENTS

Reports

59 Return to Mercury: A Global Perspective on MESSENGER's First Mercury Flyby
S. C. Solomon et al.

62 Spectroscopic Obervations of Mercury's Surface Reflectance During MESSENGER's First Mercury Flyby
W. E. McClintock et al.

66 Reflectance and Color Variations on Mercury: Regolith Processes and Compositional Heteroegneity
M. S. Robinson et al.

69 Volcanism on Mercury: Evidence from the First MESSENGER Flyby
J. W. Head et al.

73 Geology of the Caloris Basin, Mercury: A View from MESSENGER
S. L. Murchie et al.

77 Laser Altimeter Observations from MESSENGER's First Mercury Flyby
M. T. Zuber et al.

79 Mercury Cratering Record Viewed from MESSENGER's First Flyby
R. G. Strom et al.

82 The Structure of Mercury's Magnetic Field from MESSENGER's First Flyby
B. J. Anderson et al.

85 Mercury's Magnetosphere After MESSENGER's First Flyby
J. A. Slavin et al.

90 MESSENGER Observations of the Composition of Mercury's Ionized Exosphere and Plasma Environment
T. H. Zurbuchen et al.

92 Mercury's Exosphere: Observations During MESSENGER's First Mercury Flyby
W. E. McClintock et al.

See also related Science *Podcast*

Science

INTRODUCTION

Mercury, Up-Close Again

PLANETARY SCIENCE IS MUCH ABOUT COMPARISONS. STUDYING OTHER BODIES IN OUR solar system with different sizes and compositions provides essential context for understanding Earth's formation and evolution. Three planets—Mercury, Venus, and Mars—and the Moon, are most like Earth in their initial composition and relative size, but their differences are enlightening. The Moon, Venus, and especially Mars have all been visited and probed recently by spacecraft. Now it is Mercury's turn. The MESSENGER* spacecraft flew by the planet and observed it in January. The papers in this Special Issue feature these observations. After two more passes, it will settle into orbit around Mercury in 2011.

Mercury was visited by one earlier spacecraft, Mariner 10, in the mid-1970s. Its observations, and some difficult ground-based studies, provided most of our information on the planet, but raised many enigmatic questions. Despite being the smallest planet, Mercury has an actively generated magnetic field like that on Earth, but unlike that on the Moon, Mars, and Venus, and a huge iron-rich inner and molten outer core. Images of half of the surface revealed abundant scarps thought to indicate contractional faults, implying that Mercury was originally a bit larger early in its history. The extent to which its surface was shaped by impacts or volcanism was uncertain and debated. Its proximity to the Sun means that it has intense interactions with the solar wind, which, despite the magnetic field, impacts the surface of the planet, altering it and excavating some ions. Ground-based radar data hinted that ice may be present in shadowed regions of the poles.

The first MESSENGER observations provide some important early answers and a wealth of data for further study. Observations of the surface by several instruments reveal that Mercury does have a volcanic history. More contractional faults are seen in areas observed by Mariner and in an additional ~20% of the planet seen up close for the first time. A variety of impact craters help reveal relative ages of surface units and enrich the role of impact processes in shaping the planet. Impacts and solar bombardment have greatly weathered the surface; most of the iron is not in silicate minerals but apparently in nanoscale metal or oxide grains. In addition to sodium, MESSENGER detected ablation of calcium from the planet and in its magnetotail and measured in greater detail the magnetic field and its effect on its space environment.

This first pass did not cover the polar regions, so confirmation of the presence of ice will have to await future observations, which will also image other parts of the planet and come as the Sun's activity increases. But already MESSENGER has helped fill in Mercury's history and environment, allowing a better understanding of all the terrestrial planets, including Earth.

–BROOKS HANSON

*MESSENGER is an acronymn for MErcury Surface, Space ENvironment, GEochemistry, and Ranging.

CREDITS: NASA/JOHNS HOPKINS UNIVERSITY APPLIED PHYSICS LABORATORY/ARIZONA STATE UNIVERSITY/CARNEGIE INSTITUTION OF WASHINGTON

Published by AAAS

REPORT

Return to Mercury: A Global Perspective on MESSENGER's First Mercury Flyby

Sean C. Solomon,[1]* Ralph L. McNutt Jr.,[2] Thomas R. Watters,[3] David J. Lawrence,[2] William C. Feldman,[4] James W. Head,[5] Stamatios M. Krimigis,[2,6] Scott L. Murchie,[2] Roger J. Phillips,[7] James A. Slavin,[8] Maria T. Zuber[9]

In January 2008, the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft became the first probe to fly past the planet Mercury in 33 years. The encounter revealed that Mercury is a dynamic system; its liquid iron-rich outer core is coupled through a dominantly dipolar magnetic field to the surface, exosphere, and magnetosphere, all of which interact with the solar wind. MESSENGER images confirm that lobate scarps are the dominant tectonic landform and record global contraction associated with cooling of the planet. The history of contraction can be related to the history of volcanism and cratering, and the total contractional strain is at least one-third greater than inferred from Mariner 10 images. On the basis of measurements of thermal neutrons made during the flyby, the average abundance of iron in Mercury's surface material is less than 6% by weight.

Mercury, the closest planet to the Sun, is the smallest of the inner planets of our solar system and in many ways the most unusual (*1*). Its high bulk density implies that an iron-rich core makes up 60% or more of its mass (*2*), a fraction at least twice that of any other planet. Mercury's heavily cratered surface points to early cessation of internal geological activity (*3*), yet its outer core is molten (*4*), and it is the only inner planet other than Earth to have an internal magnetic field (*5*). The first spacecraft to visit Mercury was Mariner 10, which flew by three times in 1974 and 1975 and imaged about 45% of the planet's surface (*6*). In January 2008, the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft (*7*) became only the second probe to encounter the planet en route to its insertion into orbit about Mercury in March 2011. The broad range of observations made during MESSENGER's first flyby illuminate the strongly dynamic interactions among Mercury's interior, surface, tenuous atmosphere, and magnetosphere.

The January flyby was the first of three, each to be followed by a propulsive maneuver near the next aphelion, needed to reduce the arrival speed at Mercury to the point that orbit insertion can be accomplished (*8*). The Mercury flybys and subsequent maneuvers yield successive orbits having ratios of the orbital period of Mercury to that of the spacecraft of about 2:3, 3:4, and 5:6. In January, the spacecraft approached Mercury from the night side and crossed the dawn terminator shortly after closest approach at an altitude of 201 km (Fig. 1). MESSENGER viewed the sunlit side of Mercury, including about 21% of the planet's surface never imaged by Mariner 10, primarily on departure. Observations totaling about 500 MB in volume were acquired by all MESSENGER instruments, including 1213 images obtained by the Mercury Dual Imaging System (MDIS) (*9*).

Mercury's iron-rich core is central to the dynamical interactions that govern the planet's geology, exosphere, and magnetosphere. The planet's magnetic field is the result of a magnetic dynamo in the molten outer core, although whether that dynamo is currently operating (*10*–*12*) or operated only in the past and imparted a long-wavelength remanent or frozen field to Mercury's outer crust (*13*) has been a matter of debate. MESSENGER confirmed that Mercury's internal field is dominantly dipolar and indicated that there may be a quadrupole component, but no shorter-wavelength crustal anomalies were detected near closest approach (*14*). The last two results are not supportive of an entirely remanent field (*13*) and therefore point to a modern dynamo, but the contribution of crustal fields cannot be fully assessed until low-altitude measurements are made over more of the surface. Because maintenance of a dynamo requires an energy source such as freezing of an inner core or precipitation of solid iron from an outer core containing lighter elements alloyed with iron (*15*), the history of Mercury's magnetic field is closely tied to the core's thermal history and bulk composition. That core thermal history is likely expressed in the deformation of Mercury's surface.

For the 45% of Mercury's surface viewed by Mariner 10, the dominant deformational structures are lobate scarps, interpreted on the basis of morphology and the deformation of earlier impact features to be the surface expression of thrust faults formed by horizontal shortening of the crust (*16*). Lobate scarps cut across all major geological units and display a broad distribution of orientations. These characteristics led to the hypothesis that lobate scarps formed during an episode of global contraction that followed the end of heavy-impact bombardment of the inner solar system (*16*). The cumulative amount of contractional strain accommodated by the lobate scarps mapped from Mariner 10 images, inferred from topographic relief (*16*) and dimensional scaling relations (*17*), is 0.05 to 0.1%. When extrapolated to the entire planetary surface, this total strain provides an important constraint on models for the thermal history of Mercury's core and mantle (*18*). A number of questions were raised by the Mariner 10 results, however, including whether the 55% of the surface not imaged during that mission would display similarly pervasive contrac-

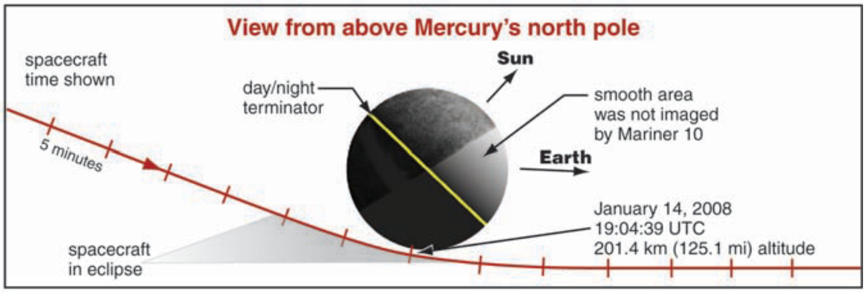


Fig. 1. Trajectory of the first MESSENGER flyby of Mercury viewed in a Mercury-fixed coordinate system from above Mercury's north pole. The spacecraft traveled from left to right. Shown are the time the spacecraft was in eclipse, the position of the terminator during the flyby, the hemisphere of Mercury previously imaged by Mariner 10, and the point of closest approach.

[1]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA. [2]Space Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [3]Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Washington, DC 20560, USA. [4]Planetary Science Institute, Tucson, AZ 85719, USA. [5]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [6]Academy of Athens, Athens 11527, Greece. [7]Southwest Research Institute, Boulder, CO 80302, USA. [8]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [9]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: scs@dtm.ciw.edu

tional faults, whether the identified scarps provide a reliable means for estimating the total contraction since the era of heavy bombardment, and whether limits could be placed on when the contraction occurred relative to other geological events preserved at Mercury's surface.

MESSENGER images of an additional 21% of Mercury's surface not previously viewed by spacecraft show that contractional fault structures are widespread and diverse in geometry. As in the area imaged by Mariner 10, lobate scarps are the most prominent tectonic landform. Such scarps range to 600 km in length (Fig. 2). Other contractional features, including wrinkle ridges and high-relief ridges, are also evident. The only areas imaged by either Mariner 10 or MESSENGER within which extensional, rather than contractional, faults have been documented are the interior of the 1550-km-diameter Caloris basin (*19*) and a small portion of the inner floor of the younger 250-km-diameter Raditladi peak-ring basin (*20*); for both features, the extensional faults are probably the result of postimpact uplift of the basin floor (*16*, *19*). MESSENGER images provide numerous examples of craters that have been substantially deformed and shortened by younger lobate scarps (Fig. 3), confirming that the scarps are contractional and providing additional opportunities to infer the magnitude of horizontal shortening accommodated in such areas. In the examples of Fig. 3, the horizontal displacement on the faults beneath each of the lobate scarps must have been at least one to several kilometers to account for the distortions of the older craters they have cut.

MESSENGER obtained images of many areas viewed by Mariner 10 but at different resolution and, at least as importantly, at different lighting conditions. Many tectonic features not recognized from Mariner 10 images can be identified in MESSENGER images of those same areas (Fig. 4). These newly recognized features indicate that the average contractional strain of Mercury's surface recorded by lobate scarps exceeds the estimates obtained from Mariner 10 observations alone. The summed length of lobate scarps in the portion of the surface imaged by both Mariner 10 and MESSENGER, together with a displacement-length scaling relation for faults on Mercury (*17*), yields an average contraction one-third greater than previous estimates. Moreover, the average contraction estimated from the total length of scarps recognized in regions newly imaged by MESSENGER is comparable to this larger figure. Because neither MESSENGER nor Mariner 10 images were obtained at optimum lighting conditions for the recognition of low-relief tectonic features in all areas, this new estimate is a minimum.

Most models of the cooling of Mercury's mantle and core (*18*) have predicted that the accumulated contractional strain since the end of heavy bombardment ~3.8 billion years ago was greater than the strain estimated from the geometry of lobate scarps identified in Mariner 10 images (*17*). The models most consistent with the Mariner 10 results had a comparatively creep-resistant (anhydrous) mantle, slowly decaying interior heat production (dominated by ^{232}Th as opposed to the shorter-lived ^{235}U and ^{40}K), and a large amount (>6% by weight) of a lighter element such as S in Mercury's outer core to retard the growth of a solid inner core (*18*). An increase by at least one-

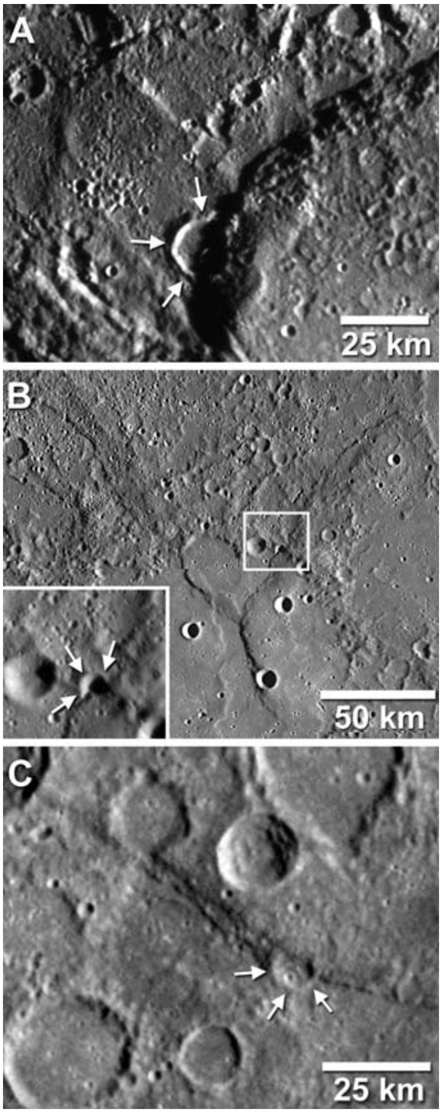


Fig. 3. Three examples of craters substantially deformed by a lobate scarp. In each case, portions of the crater floor and rim have been buried by overthrusted material. (**A**) The northern segment of Beagle Rupes has crosscut a ~17-km-diameter impact crater (centered near 0.3°N, 101°E; arrows) on intercrater plains. From NAC frame EN0108827037M. (**B**) The northeast-southwest–trending segment of a lobate scarp has cut a ~5-km-diameter impact crater (centered near 7.9°S, 108.2°E; see inset) located near the rim of a larger degraded impact crater that was flooded by smooth plains and subsequently deformed by wrinkle ridges. This mosaic consists of NAC frames EN0108825899M, EN0108825904M, EN0108825994M, and EN0108825999M. (**C**) A northwest-southeast–trending lobate scarp has deformed an ~11-km-diameter crater (centered near 16.5°S, 133°E) on intercrater plains. From NAC frame EN0108828317M.

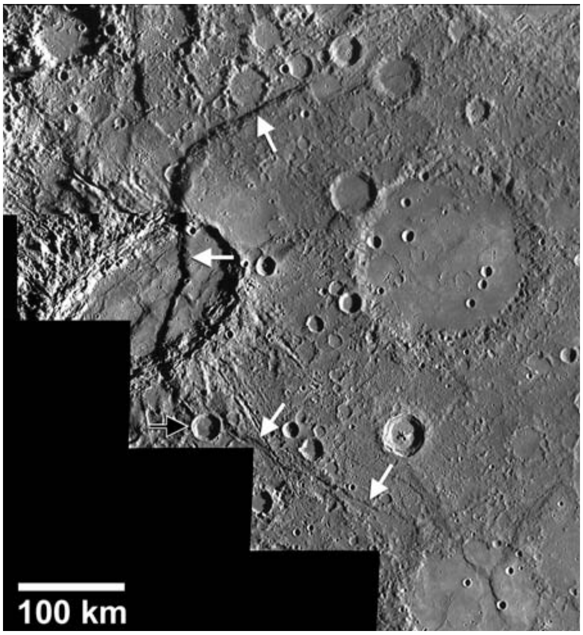


Fig. 2. Beagle Rupes, a prominent lobate scarp (white arrows) imaged on the portion of Mercury's surface viewed for the first time by MESSENGER, is more than 600 km long and offsets the floor and walls of the ~220-km-diameter, elliptically shaped impact crater Sveinsdóttir. The floor of the impact crater was flooded by smooth plains and subsequently deformed by wrinkle ridges before scarp development. Beagle Rupes is one of the most arcuate of the lobate scarps found on Mercury to date. A ~30-km-diameter crater sits undeformed on the northwest-southeast segment of Beagle Rupes (black arrow). This MDIS monochrome (750-nm) narrow-angle camera (NAC) image mosaic is centered at about 3°S, 103.5°E; north is to the top in this and other images and mosaics. The relative positions on the planet of the area in this image, other images in this paper, and images and profiles in companion papers are depicted in fig. S1. The mosaic consists of NAC frames EN0108825899M, EN0108825904M, EN0108826004M, EN0108826095M, EN0108826100M, EN0108826105M, EN0108826191M, EN0108826196M, EN0108826201M, EN0108826206M, EN0108827037M, and EN0108827042M.

third in the average preserved contractional strain will relax one or more of these model restrictions and permit a greater range of possible planetary thermal histories.

Additional information on the timing of global contraction relative to other major events in Mercury's geological evolution can be derived from MESSENGER observations. From Mariner 10 images it was seen that scarps deform all major geological units, including the comparatively young smooth plains, but no instance of a scarp embayed by plains was recognized (*16*). A candidate for such an embayment relation was imaged by MESSENGER (Fig. 5). This and similar relations, together with the abundance of evidence that smooth plains are volcanic deposits (*19*, *21*), indicate that scarp development began before many smooth plains were emplaced and continued after the eruption of the youngest appreciable expanse of smooth plains material yet observed. Whereas a number of lobate scarps deformed older craters (Figs. 2 to 4), there are also many examples of undeformed craters superposed on scarps (e.g., Figs. 2 and 5). These relations offer the promise that the rate of global contraction subsequent to late heavy bombardment can be estimated and tied to the history of plains emplacement. Such records would constrain the evolution of mantle temperatures and the rate of growth of the solid inner core and its potential as a power source for Mercury's core dynamo.

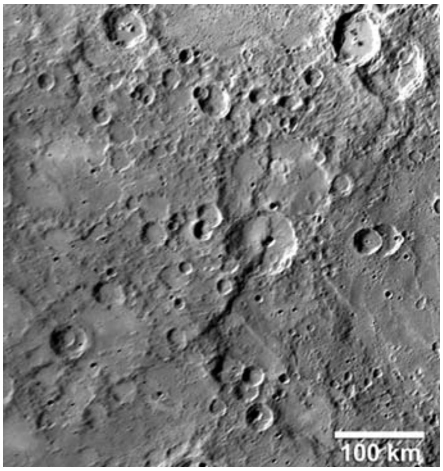


Fig. 4. This ~270-km-long lobate scarp, which deformed the two large craters in the center of the mosaic, was not visible in Mariner 10 images of the area, because during the Mariner 10 flybys the Sun was locally at a high angle to the surface. This mosaic is centered near 24°S, 254°E, and uses images acquired before the flyby closest approach. The mosaic consists of NAC frames EN0108821370M, EN0108821375M, EN0108821397M, and EN0108821402M.

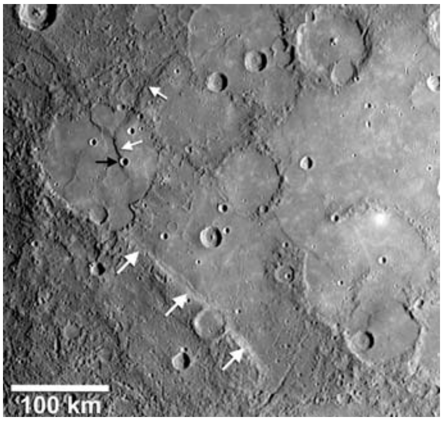


Fig. 5. Possible embayment of lobate scarps by smooth plains. Smooth plains material, interpreted to consist of volcanic flows, appears to have ponded against the structural relief of a preexisting lobate scarp formed in older intercrater plains (lower set of thick white arrows). Low-relief ridges in the smooth plains just outward of the scarp face may be evidence of continued movement on the underlying thrust fault after plains emplacement. In the upper left (upper set of thin white arrows), a lobate scarp cuts across both intercrater plains (topmost arrow) and smooth plains that filled the floor of a ~120-km-diameter impact crater (Fig. 3B). Slip on the fault scarp appears to postdate the formation of wrinkle ridges formed in the smooth plains interior to the crater. Shown also is the undeformed crater superposed on this scarp (black arrow). This mosaic is centered near 10°S, 110°E, and uses images acquired on departure from closest approach. The mosaic consists of NAC frames EN0108828307M, EN0108828312M, EN0108828359M, and EN0108828364M.

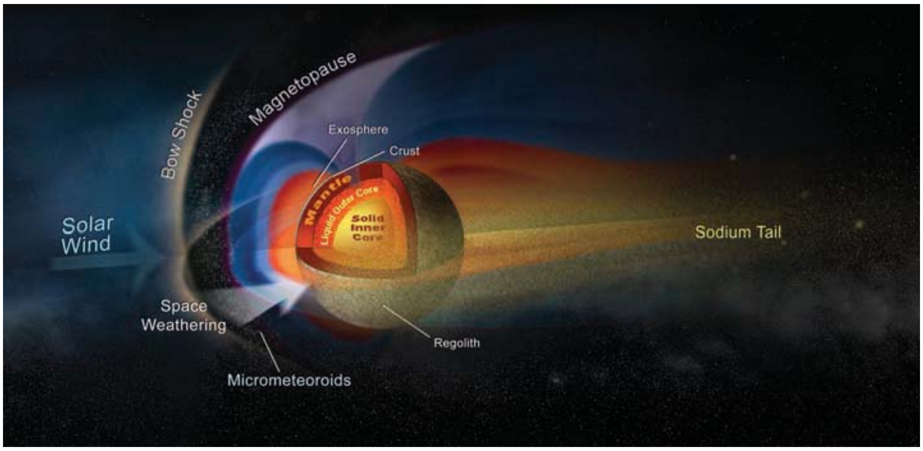


Fig. 6. Schematic depiction of several of the interconnections among Mercury's interior, surface, exosphere, magnetosphere, and interplanetary environment.

Even though Mercury is more than 60% Fe by weight, the average Fe abundance of Mercury's surface materials, and by inference its crust and mantle, is lower than those of the other inner planets (*22*). This contrast is rooted in planetary formational processes (*1*), but distinguishing among competing hypotheses requires accurate measurements of the structure (*23*) and major-element chemistry (*24*) of Mercury's crust. MESSENGER detected no absorption features attributable to Fe^{2+} in silicates either in disk-averaged or higher–spatial resolution visible and near-infrared spectra (*25*) or with multispectral imaging (*26*). The generally red spectral slopes displayed by Mercury surface materials (*25*, *26*) have been attributed to nanometer-scale particles of Fe metal, originating from meteoritic iron or reduction of iron-bearing surface minerals and redeposited from vapor by space weathering processes accompanying meteoroid and charged-particle impacts (*27*). The most direct information on surface Fe abundance must await measurements by the Gamma-Ray and Neutron Spectrometer (GRNS) (*28*) once MESSENGER is in orbit about Mercury.

An upper limit on surface Fe abundance can be estimated from measurements made by the Neutron Spectrometer (NS) sensor on GRNS. Thermal neutrons provide information about surface abundances of neutron-absorbing elements, e.g., Fe, Ti, Gd, and Sm (*29*). MESSENGER's NS can measure thermal neutrons with Doppler filter spectroscopy (DFS) (*30*), which uses the spacecraft speed (~7 km/s) to separate slowly moving thermal neutrons (~2 km/s) from more energetic epithermal neutrons. DFS was applied to the flyby observations and made use of a 90° spacecraft rotation near closest approach to provide separate measurements of Doppler shifted (J_+) and nonshifted (J_0) neutrons. The ratio $J_0/(J_+ - J_0)$ provides an estimate of the thermal neutrons from Mercury, which can then be related to the abundances of neutron-absorbing elements. Lunar soils provide approximate analogs to Mercury surface materials because of their low H content and their wide ranges of Fe and Ti abundances. On the basis of a comparison of Mercury flyby measurements with modeled neutron fluxes for a range of lunar soils (*31*), the upper-limit neutron absorption content is less than that of the comparatively low-Fe Luna 20 soil, for which neutron absorption is nonetheless dominated by Fe, at 5.8% by weight (*32*). If all neutron absorption in Mercury's soil were due to Fe, then the MESSENGER data suggest that the

Fe abundance is less than ~6% and would be lower still if Ti, Gd, or Sm is present. For comparison, the average Fe abundance is about 5% and 8% in Earth's continental and oceanic crust, respectively, and 5% in lunar highlands crust (*33*).

During its flyby, MESSENGER provided a comprehensive view of solar wind interaction with Mercury's magnetic field and neutral atmosphere and, indirectly, its surface. At the time of the flyby, solar activity was low and, in contrast with Mariner 10 observations, no energetic charged particles with energies above ~30 keV were detected (*34*). Magnetometer observations (*14*) of the magnetospheric boundaries, current systems, and plasma waves confirm that this magnetosphere appears structurally to be a miniature of that of Earth. One notable difference is the presence of a double current sheet at the dawn terminator that likely represents heavy planetary ion effects unique to Mercury (*34*). The Mercury Atmospheric and Surface Composition Spectrometer observed neutral Na and Ca in Mercury's exosphere—delivered from surface materials in part by the same micrometeoroid and ion-impact processes that space-weather the surface—and mapped the structure of Mercury's antisunward Na tail (*35*). The Fast Imaging Plasma Spectrometer sensor on the Energetic Particle and Plasma Spectrometer instrument (*36*) observed a range of heavy magnetospheric plasma ions—including O^+, Na^+, Mg^+, K^+, Ca^+, S^+, and H_2S^+—derived from the exosphere or surface (*37*). On the basis of the full set of observations made during MESSENGER's first flyby, Mercury is seen to be a dynamic planet where the interactions among core, surface, exosphere, magnetosphere, and interplanetary environment are strongly interlinked (Fig. 6). Subsequent encounters under different solar conditions and one Earth year in orbit about Mercury as the Sun approaches the next maximum in the solar cycle should permit MESSENGER to explore these interactions across their full range of behavior.

References and Notes

1. S. C. Solomon, *Earth Planet. Sci. Lett.* **216**, 441 (2003).
2. H. Harder, G. Schubert, *Icarus* **151**, 118 (2001).
3. P. D. Spudis, J. E. Guest, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, 1988), pp. 118–164.
4. J. L. Margot, S. J. Peale, R. F. Jurgens, M. A. Slade, I. V. Holin, *Science* **316**, 710 (2007).
5. N. F. Ness, K. W. Behannon, R. P. Lepping, Y. C. Whang, *J. Geophys. Res.* **80**, 2708 (1975).
6. B. C. Murray, *J. Geophys. Res.* **80**, 2342 (1975).
7. S. C. Solomon, R. L. McNutt Jr., R. E. Gold, D. L. Domingue, *Space Sci. Rev.* **131**, 3 (2007).
8. J. V. McAdams, R. W. Farquhar, A. H. Taylor, B. G. Williams, *Space Sci. Rev.* **131**, 219 (2007).
9. S. E. Hawkins III *et al.*, *Space Sci. Rev.* **131**, 247 (2007).
10. S. Stanley, J. Bloxham, W. E. Hutchison, M. T. Zuber, *Earth Planet. Sci. Lett.* **234**, 27 (2005).
11. M. H. Heimpel, J. M. Aurnou, F. M. Al-Shamali, N. Gomez Perez, *Earth Planet. Sci. Lett.* **236**, 542 (2005).
12. U. R. Christensen, *Nature* **444**, 1056 (2006).
13. O. Aharonson, M. T. Zuber, S. C. Solomon, *Earth Planet. Sci. Lett.* **218**, 261 (2004).
14. B. J. Anderson *et al.*, *Science* **321**, 82 (2008).
15. B. Chen, J. Li, S. A. Hauck II, *Geophys. Res. Lett.* **35**, L07201, 10.1029/2008GL033311 (2008).
16. R. G. Strom, N. J. Trask, J. E. Guest, *J. Geophys. Res.* **80**, 2478 (1975).
17. T. R. Watters, M. S. Robinson, A. C. Cook, *Geology* **26**, 991 (1998).
18. S. A. Hauck II, A. J. Dombard, R. J. Phillips, S. C. Solomon, *Earth Planet. Sci. Lett.* **222**, 713 (2004).
19. S. L. Murchie *et al.*, *Science* **321**, 73 (2008).
20. R. G. Strom, C. R. Chapman, W. J. Merline, S. C. Solomon, J. W. Head III, *Science* **321**, 79 (2008).
21. J. W. Head *et al.*, *Science* **321**, 69 (2008).
22. M. S. Robinson, G. J. Taylor, *Meteorit. Planet. Sci.* **36**, 841 (2001).
23. M. T. Zuber *et al.*, *Science* **321**, 77 (2008).
24. W. V. Boynton *et al.*, *Space Sci. Rev.* **131**, 85 (2007).
25. W. E. McClintock *et al.*, *Science* **321**, 62 (2008).
26. M. S. Robinson *et al.*, *Science* **321**, 66 (2008).
27. B. Hapke, *J. Geophys. Res.* **106**, 10039 (2001).
28. J. O. Goldsten *et al.*, *Space Sci. Rev.* **131**, 339 (2007).
29. W. C. Feldman *et al.*, *J. Geophys. Res.* **105**, 20347 (2000).
30. W. C. Feldman, D. M. Drake, *Nucl. Instrum. Methods Phys. Res. A* **245**, 182 (1986).
31. D. J. Lawrence *et al.*, *J. Geophys. Res.* **111**, E08001, 10.1029/2005JE002637 (2006).
32. L. Haskin, P. Warren, in *Lunar Sourcebook*, G. Heiken, D. Vaniman, B. M. French, Eds. (Cambridge Univ. Press, New York, 1991), pp. 357–474.
33. K. Lodders, B. Fegley Jr., *The Planetary Scientist's Companion* (Oxford Univ. Press, New York, 1998), pp. 140 and 177.
34. J. A. Slavin *et al.*, *Science* **321**, 85 (2008).
35. W. E. McClintock *et al.*, *Science* **321**, 92 (2008).
36. G. B. Andrews *et al.*, *Space Sci. Rev.* **131**, 523 (2007).
37. T. H. Zurbuchen *et al.*, *Science* **321**, 90 (2008).
38. The hundreds of engineers and technical support personnel who brought MESSENGER from a concept to a successful flight project warrant the sustained appreciation of the mission science team. We also thank L. M. Prockter for assembling fig. S1. The MESSENGER project is supported by the NASA Discovery Program under contracts NASW-00002 to the Carnegie Institution of Washington and NAS5-97271 to Johns Hopkins University Applied Physics Laboratory.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/59/DC1
Fig. S1
References

28 April 2008; accepted 3 June 2008
10.1126/science.1159706

REPORT

Spectroscopic Observations of Mercury's Surface Reflectance During MESSENGER's First Mercury Flyby

William E. McClintock,[1]* Noam R. Izenberg,[2] Gregory M. Holsclaw,[1] David T. Blewett,[2] Deborah L. Domingue,[2] James W. Head III,[3] Jörn Helbert,[4] Timothy J. McCoy,[5] Scott L. Murchie,[2] Mark S. Robinson,[6] Sean C. Solomon,[7] Ann L. Sprague,[8] Faith Vilas[9]

During MESSENGER's first flyby of Mercury, the Mercury Atmospheric and Surface Composition Spectrometer made simultaneous mid-ultraviolet to near-infrared (wavelengths of 200 to 1300 nanometers) reflectance observations of the surface. An ultraviolet absorption (<280 nanometers) suggests that the ferrous oxide (Fe^{2+}) content of silicates in average surface material is low (less than 2 to 3 weight percent). This result is supported by the lack of a detectable 1-micrometer Fe^{2+} absorption band in high-spatial-resolution spectra of mature surface materials as well as immature crater ejecta, which suggests that the ferrous iron content may be low both on the surface and at depth. Differences in absorption features and slope among the spectra are evidence for variations in composition and regolith maturation of Mercury's surface.

Before MESSENGER's first flyby of Mercury on 14 January 2008, our knowledge of the planet's surface mineralogy came from low-spatial-resolution, ground-based spectroscopic observations. Early disk-integrated observations showed a low-albedo, relatively featureless spectrum that increases monotonically across the visible to near-infrared wavelengths, a characteristic referred to as a "red" spectral slope (*1*–*3*). On the Moon these spectral characteristics result from space weathering, a process in which amorphous silica coatings containing nanometer-scale metallic iron (nanophase metallic iron, $npFe^0$) darken exposed regolith, increase spectral slope, and reduce spectral contrast (*4*, *5*). Although Mercury's surface appears to be highly space-weathered, the absence of identifiable near-infrared absorptions argues for a low average ferrous iron content (*2*, *6*, *7*). This view is supported by recent mid-infrared emission spectra and reflectance observations that have been interpreted to indicate the presence of plagioclase feldspar and low-iron orthopyroxene (enstatite) (*6*, *8*, *9*) and high-Ca clinopyroxene (*10*). These mid-infrared spectra support the presence of Na-bearing feldspar and Mg-rich minerals (*8*), which are present in very-low-iron terrestrial rock types. To more fully understand Mercury's surface composition, it is necessary to explore regional spectral variations across units that contain relatively unweathered materials, such as ejecta from small craters, which are often less than 100 km in diameter. This has not been possible from Earth because long atmospheric path lengths and low spatial resolution (>200 km) hinder the observations.

During MESSENGER's flyby, the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) (*11*) measured high-spatial-resolution reflectance spectra from Mercury's surface. MASCS consists of a small Cassegrain telescope that simultaneously feeds a Visible and Infrared Spectrograph (VIRS) and an Ultraviolet and Visible Spectrometer (UVVS), covering the wavelengths 325 to 1300 nm and 220 to 320 nm, respectively. Both VIRS (0.023° circular field of view) and UVVS (0.04° × 0.05° rectangular field of view) are point instruments whose fields of view (FOVs) are offset by 0.38° in the telescope focal plane. The MESSENGER observations constitute the first reflectance measurements at high spatial and high spectral resolution, as well as the first middle ultraviolet (MUV) spectra (220 to 350 nm), of Mercury's surface.

Nine days after closest approach, MASCS obtained full-disk spectra of Mercury at a phase angle of 87° and a planetary diameter of 0.068°. In Fig. 1A, these are compared with spectra of the southern hemisphere of the lunar farside—which contains feldspathic highlands, the interior of the South Pole–Aitken Basin, and some maria—obtained by MASCS on 31 July 2005. Observations using a single spacecraft instrument and similar viewing geometry allow us to compare directly the reflectance of Mercury's surface to that of the Moon, which has been extensively studied using both remote sensing techniques and laboratory analysis of soil and rock samples acquired during the Apollo and Luna programs. To facilitate direct comparison, we used an empirical multispectral phase function of the Moon derived from Robotic Lunar Observatory (ROLO) observations (*12*) to adjust the MASCS lunar spectrum to 87° phase angle.

The MASCS disk-averaged spectra of Mercury and the Moon display the characteristic red slope at visible wavelengths (350 to 600 nm) observed from Earth (*3*, *12*); however, Mercury's visible spectral slope is 10% less than that of the Moon. The shallower slope continues into the infrared, but the difference becomes less pronounced (*13*). Throughout the MUV, the lunar spectrum obtained by MASCS is also observed to increase in a nearly linear fashion (Fig. 1B). This linear trend is present in spectra of the lunar nearside obtained from Earth orbit (*14*, *15*) and in laboratory spectra of returned lunar soil samples (*16*). In contrast, Mercury's disk-averaged MUV (<300 nm) reflectance displays an abrupt downturn toward shorter wavelengths relative to the linear-sloped visible trend. This downturn is also observed in the spatially resolved UV spectrum of a crater filled with smooth plains material (location A in Fig. 2).

Laboratory spectra of lunar rock powders exhibit strong trough-like features (*16*) that result from the presence of oxygen–transition metal charge transfer (OMCT) bands, which are the main cause for absorption in silicate materials at wavelengths of 200 to 400 nm (*17*, *18*). Trough-like features are not observed in UV spectra of the lunar nearside taken from Earth orbit (*14*, *15*) or in laboratory spectra of both mature and immature lunar soils (*16*) because of space weathering by micrometeorite impacts and solar wind sputtering. A primary mechanism causing suppression of UV bands is the reduction of ferrous iron followed by the vapor deposition of a npFe^0-rich patina (*4*, *19–21*). This patina masks the UV absorption within the underlying silicate material and erodes the long-wavelength absorption edge, producing a spectrum with a nearly linear slope extending to visible wavelengths (*4*, *22*), and substantially reduces the contrast observed in Fe^{2+} bands that occur in the infrared near 1 μm.

The location of the downturn in Mercury's spectrum, with its edge near 280 nm (>20% depression below the linear continuum), and its troughlike long-wavelength shoulder are consistent with the presence of OMCT bands caused by Fe^{2+} and/or Ti^{4+}, which have band centers near 250 nm (*18*, *23*). Furthermore, the depth and shape of the band argue that the ferrous iron content is low in the average surface materials. Laboratory UV reflectance spectra show that the inclusion of even 2 to 3 weight percent Fe^{2+} in silicates (*24*) and glasses (*25*) results in a deep central absorption and a movement of the absorption edge from 270 nm toward 300 nm. An OMCT feature due to Fe^{2+} can be present at 250 nm even if iron abundance were too low to produce a detectable

[1]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA. [2]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [3]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [4]Institute of Planetary Research, Deutsches Zentrum für Luft- und Raumfahrt, Berlin 12489, Germany. [5]National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA. [6]Department of Geological Sciences, Arizona State University, Tempe, AZ 85287, USA. [7]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA. [8]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [9]MMT Observatory, University of Arizona, Tucson, AZ 85721, USA

*To whom correspondence should be addressed. E-mail: william.mcclintock@lasp.colorado.edu

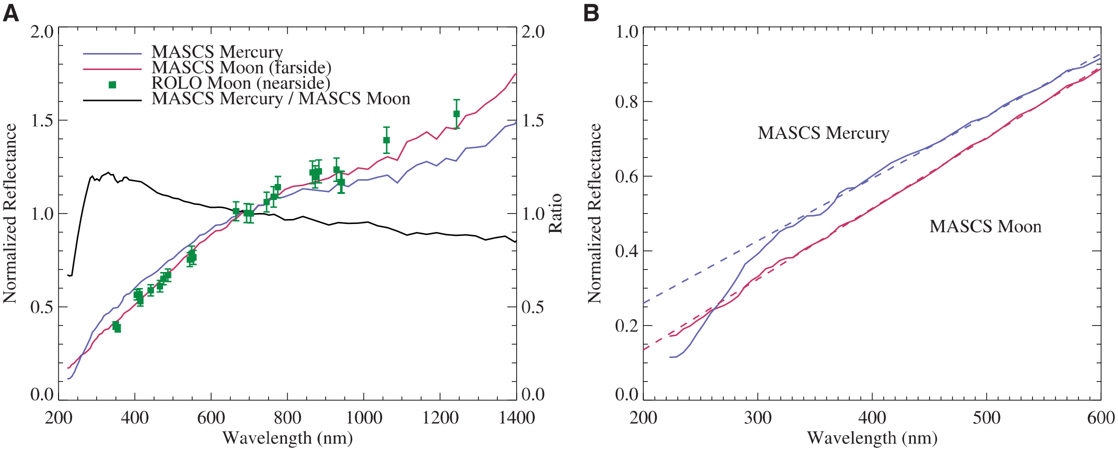


Fig. 1. (**A**) MASCS disk-averaged reflectance (observed spectral radiance of the surface divided by the radiance from a hypothetical normally illuminated Lambertian disk) of Mercury and the Moon, along with a ROLO-derived full-disk spectrum of the waxing phase (87°) of the lunar nearside, which is predominantly composed of highlands terrain. These spectra are scaled to a value of 1.0 at 700 nm. The excellent agreement between the adjusted MASCS lunar observations and the ROLO spectrum provides confidence in the instrument radiometric performance and a reference for interpreting the Mercury data. The ratio of Mercury reflectance to that of the Moon highlights the differences between the spectral properties of the two bodies. (**B**) Mercury's MUV reflectance (blue curve) exhibits a distinct departure from a linear trend (blue dashed line) at wavelengths <300 nm that is not observed for the Moon (red curved and dashed line).

band near 1 μm, because the OMCT band arises from a strong, allowed transition, whereas the near-infrared band is a forbidden (and thus weak) transition (*17*). Assemblages containing iron-poor pyroxene or olivine, or plagioclase feldspar with small amounts of iron, produce laboratory UV absorption spectra consistent with the current data (*24*). These mineralogical interpretations are similar to those made from ground-based mid-infrared spectra of Mercury (*26*). Ti^{4+} could also produce an OMCT feature near 250 nm without exhibiting a corresponding 1-μm band. Because Fe^{2+} abundance in silicates appears to be very low for the surfaces observed by MASCS, nonsilicate phases such as iron-bearing opaque minerals (*27*) and sulfides or meteoritic infall (*28*) could provide the source of the nanometer-scale metallic particles that darken and redden Mercury's spectrum.

If the vapor-deposited metallic particles are <25 nm in diameter, then the presence of a trough-like OMCT feature in Mercury's spectrum limits the amount of npFe^0 (or possibly nanometer-scale Ti^0 particles, which have virtually identical optical absorption properties) in the patinas and coatings of surface materials to 0.1 to 0.2 weight percent, because larger amounts produce broad UV absorptions that would erode the shape of bands (*4*). Laboratory experiments with regolith analogs indicate that larger particle size tends to lower reflectance throughout the visible and near-infrared with little change in overall shape (*22*). The size range for metallic iron particles that contribute to space weathering effects in lunar soils may vary from a few nanometers to hundreds of nanometers, but npFe^0 size on Mercury may be different. More intense micrometeoritic bombardment is expected at Mercury than at the Moon (*29*). Furthermore, Ostwald ripening—a process by which npFe^0 particles within a glass matrix coarsen at elevated temperature—could also be operating on Mercury (*28*). Recent analysis of Mariner 10 (*30*) and MESSENGER (*31*) multispectral images provides evidence that relatively coarse-grained opaque phases (i.e., particle diameters

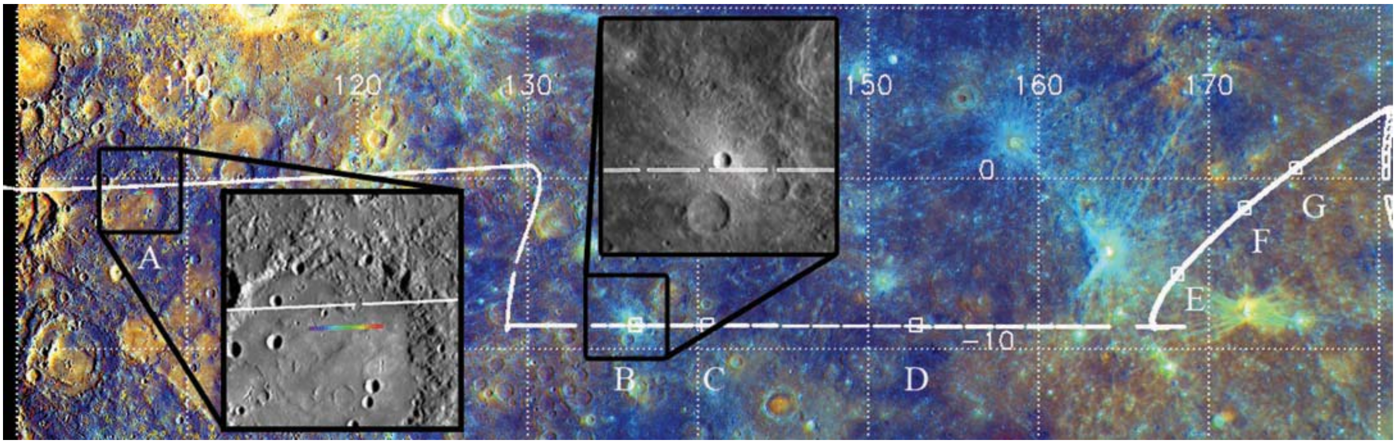


Fig. 2. A segment of the VIRS FOV track is shown in white over an MDIS image mosaic, formed by combining a high-resolution narrow-angle camera (NAC) image mosaic with a lower-resolution color image composite (*31*). MDIS images reveal a surface populated by three major color units: higher-reflectance, relatively red smooth plains; lower-reflectance material with lesser spectral slope; and intermediate material. The first part of the track crosses primarily intermediate terrains before briefly entering the northern edge of a relatively young smooth plains unit (near G) identified in MDIS color images. Locations B to G along the track correspond to the spectra shown in Fig. 3. The spatial-spectral sampling of UVVS is denoted by a rainbow for the single spectrum taken during the flyby.

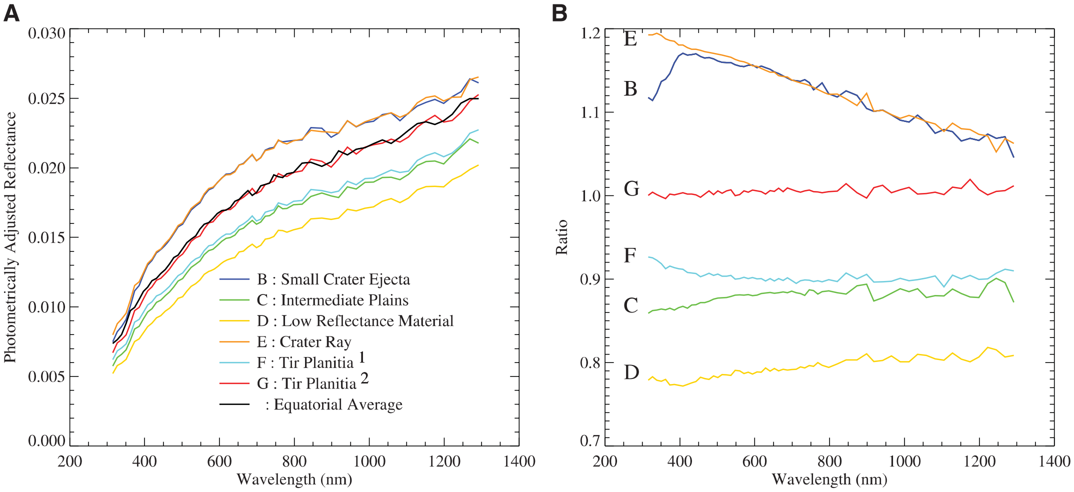


Fig. 3. (**A**) Spectra B and C are from the bright ejecta of an impact crater ~10 km in diameter (8.5°S, 136.3°E) and from an intermediate plains unit just east of the crater, respectively (see Fig. 2). Spectrum D is from low-reflectance material (8.6°S, 154.7°E). Spectrum E is from an area where a bright crater ray overlays a smooth plains unit (5.6°S, 168.2°E). Spectra from plains units (F and G) adjacent to and within Tir Planitia (3°S, 177°E) are also shown. All were adjusted to a common viewing geometry (i = 73° and e = 14.5°) and plotted on a common vertical scale. Thus, differences in magnitude and shape represent real variation in surface reflectance properties from location to location. (**B**) Ratios of the spectra shown in (A) to the average MASCS surface spectrum highlight spectral differences.

greater than that of npFe^0) play an important role in albedo and color variations on Mercury. The low, featureless reflectance typical of opaque minerals, such as ilmenite, is known to decrease the overall reflectance of a mixture and decrease its spectral slope (*27*), so a larger abundance of an opaque phase in Mercury's surface relative to that on the Moon could explain the low reflectance and the visible-infrared spectral slope.

Spatially resolved MASCS observations began near 0°N, 95°E, as the spacecraft crossed the morning terminator and continued to about 10°S, 180°E (Fig. 2). During this traverse, the FOVs primarily crossed terrains identified in Mercury Dual Imaging System (MDIS) color images (*31*) as units with intermediate reflectance, but they also sampled some low-reflectance units. The combination of spacecraft motion and changing distance to the surface caused the VIRS spectral footprint to vary from 1 × 5 km to 2 × 75 km (caused by a spacecraft slew) and back to 2 × 5 km along the ground track. All of the spectra were acquired at phase angles of ~90°, with incidence angle (i) beginning at 80° and ending at 10°. Operational constraints limited the UVVS observations to a single usable spectrum, with footprint dimensions of 1 km × 40 km, acquired while the instrument FOV crossed a smooth-floored crater near 1°S, 107°E (i = 78°). During each integration period, VIRS simultaneously records the magnitude of reflected light across all wavelengths within its spatial FOV; in contrast, UVVS builds a spectrum by scanning through wavelengths as the spatial FOV moves across the surface. The VIRS and UVVS FOV tracks are offset by 16 km at this location.

Example spectra (Fig. 3A), adjusted to a common viewing geometry by means of a Hapke photometric model (*32*), provide insight into regolith compositional differences and maturation processes on Mercury. Ratios of these spectra to an average of 492 spectra with $i < 75°$ and emission angle $e < 75°$ (Fig. 3B) highlight subtle spectral variations. Compositional variation on small scales is evidenced in the near-UV portion of the spectrum. In relatively fresh materials, the striking difference between the small crater ejecta spectrum (B) and the crater ray spectrum (E) below 400 nm (Fig. 3B) likely reflects a compositional difference between the two locations and is not solely an effect of differential maturity. Slope and curvature differences in both the visible and near-infrared among the spectra of surfaces of roughly similar maturity (C, D, F, and G, Fig. 3B) are also suggestive of compositional differences. In particular, the upturn seen below 500 nm in the low-reflectance material (D) is consistent with the presence of ferrous iron- and titanium-bearing opaque phases at that location (*31*), as seen in the spectra of some opaque-rich returned lunar samples (*16*). An equally suggestive, more pronounced, upturn is also seen in the spectrum at F—which appears in MDIS images to be a mixture of low-reflectance and intermediate materials—but is conspicuously absent in the nearby spectrum at G, which is primarily composed of intermediate material. Because spectral slope and band depth change not only with composition but also with soil grain size, detailed radiative transfer modeling will be needed to constrain the specific mineral phases present.

Lunar-like maturity trends are also clearly evident in the MASCS data. The small crater (B, Figs. 2 and 3A) and the bright ray (E, Figs. 2 and 3A) represent stratigraphically young, fresh material. Their relative spectra have negative slopes and relatively high reflectance, whereas the older mature units have a positive slope and low reflectance, commensurate with the darkening and reddening associated with space weathering. The small crater (B) shows an enhanced absorption at wavelengths less than 400 nm relative to both the adjacent plains unit (C) and the average equatorial terrains. Because the crater and the adjacent plains are likely to have the same composition, this difference could indicate that the OMCT UV band in the more mature plains material has been more obscured by space weathering, causing spectra B and C to differ in the same way as spectra for a powdered lunar rock differ from those of a lunar soil (as described earlier). This discovery indicates that the spectral shape in the near-UV may be a useful indicator of optical maturity. None of the VIRS spatially resolved spectra exhibit a 1-μm absorption band (*33*), including the freshest material. Therefore, it is very unlikely that the ferrous iron absorption on Mercury has been hidden or erased by intense space weathering (*29*). The absence of this band is further evidence that ferrous iron is only minimally present in silicates, both on the surface and at depth within the crust, in the regions sampled by MASCS.

References and Notes

1. T. B. McCord, J. B. Adams, *Icarus* **17**, 585 (1972).
2. F. Vilas, *Icarus* **64**, 133 (1985).
3. F. Vilas, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 59–76.
4. B. Hapke, *J. Geophys. Res.* **106**, 10039 (2001).
5. E. M. Fischer, C. M. Pieters, *Icarus* **111**, 475 (1994).
6. B. Hapke, G. E. Danielson Jr., K. Klaasen, L. Wilson, *J. Geophys. Res.* **80**, 2431 (1975).
7. J. Warell, *Icarus* **161**, 199 (2003).
8. A. L. Sprague, T. L. Roush, *Icarus* **133**, 174 (1998).
9. J. Warell, D. T. Blewett, *Icarus* **168**, 257 (2004).
10. J. Warell, A. L. Sprague, J. P. Emery, R. W. H. Kozlowski, A. Long, *Icarus* **180**, 281 (2006).
11. W. E. McClintock, M. R. Lankton, *Space Sci. Rev.* **131**, 481 (2007).
12. H. H. Kieffer, T. C. Stone, *Astron. J.* **129**, 2887 (2005).
13. Prior comparisons of the slope of ground-based Mercury spectra (expressed as the ratio of reflectance at 415 nm to that at 750 nm) to that of the Moon (*34*) did not find a consistent relation. We argue that the observations reported here, made with the same instrument and free from the complicating effects of the terrestrial atmosphere and stellar calibrations, are more reliable.
14. S. Janz, E. Hilsenrath, R. Cebula, T. Kelly, *Geophys. Res. Lett.* **23**, 2297 (1996).
15. G. K. Fox *et al.*, *Mon. Not. R. Astron. Soc.* **298**, 303 (1998).
16. J. K. Wagner, B. W. Hapke, E. N. Wells, *Icarus* **69**, 14 (1987).
17. R. G. Burns, in *Remote Geochemical Analysis*, C. Pieters, P. Englert, Eds. (Cambridge Univ. Press, Cambridge, 1993), pp. 3–29.
18. J. A. Tossell, D. J. Vaughn, K. H. Johnson, *Am. Mineral.* **59**, 319 (1974).
19. B. Hapke, W. Cassidy, E. Wells, *Moon* **13**, 339 (1975).
20. B. Hapke, *Phys. Earth Planet. Inter.* **15**, 264 (1977).
21. C. M. Pieters *et al.*, *Meteorit. Planet. Sci.* **35**, 1101 (2000).
22. S. K. Noble, C. M. Pieters, L. P. Keller, *Icarus* **192**, 629 (2007).
23. B. M. Loeffler *et al.*, *Geochim. Cosmochim. Acta* **3** (suppl. 5), 3007 (1974).
24. E. A. Cloutis *et al.*, *Icarus*, 10.1016/j.icarus.2008.04.018 (2008).
25. E. Wells, B. Hapke, *Science* **195**, 977 (1977).
26. A. L. Sprague *et al.*, *Meteorit. Planet. Sci.* **37**, 1255 (2002).
27. B. Rava, B. Hapke, *Icarus* **71**, 397 (1987).
28. S. K. Noble, C. M. Pieters, *Astron. Vestnik* **37**, 34 (2003) [English version in *Solar Syst. Res.* **37**, 31 (2003)].
29. M. J. Cintala, *J. Geophys. Res.* **97**, 947 (1992).
30. B. W. Denevi, M. S. Robinson, *Icarus*, 10.1016/j.icarus.2008.04.021 (2008).
31. M. S. Robinson *et al.*, *Science* **321**, 66 (2008).
32. D. L. Domingue *et al.*, *Lunar Planet. Sci.* **39**, abstract 1298 (2008).
33. Some Earth-based reflectance spectra of Mercury show possible evidence of an absorption near 1 μm (1000 nm) attributable to small amounts of ferrous iron in silicates (*3*, *10*, *35*, *36*), whereas others are remarkably featureless (*3*, *9*). Some Earth-based measurements may be affected by interference from Earth's atmosphere.
34. D. T. Blewett, P. G. Lucey, B. R. Hawke, G. G. Ling, M. S. Robinson, *Icarus* **129**, 217 (1997).
35. T. B. McCord, J. B. Adams, *Science* **178**, 745 (1972).
36. T. B. McCord, R. N. Clark, *J. Geophys. Res.* **84**, 7664 (1979).

37. We thank the NASA MESSENGER mission, instrument, planning, and operations teams at the Johns Hopkins University Applied Physics Laboratory and the University of Colorado Laboratory for Atmospheric and Space Physics for support. R. L. McNutt Jr. assisted with mission planning and M. R. Lankton led instrument design. M. Kochte, H. Kang, R. Vaughan, K. Wittenburg, R. Shelton, and A. Berman designed the instrument sequences. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to the Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington.

1 May 2008; accepted 4 June 2008
10.1126/science.1159933

REPORT

Reflectance and Color Variations on Mercury: Regolith Processes and Compositional Heterogeneity

Mark S. Robinson,[1]* Scott L. Murchie,[2] David T. Blewett,[2] Deborah L. Domingue,[2]
S. Edward Hawkins III,[2] James W. Head,[3] Gregory M. Holsclaw,[4] William E. McClintock,[4]
Timothy J. McCoy,[5] Ralph L. McNutt Jr.,[2] Louise M. Prockter,[2]
Sean C. Solomon,[6] Thomas R. Watters[7]

Multispectral images of Mercury obtained by the MESSENGER spacecraft reveal that its surface has an overall relatively low reflectance with three large-scale units identified on the basis of reflectance and slope (0.4 to 1.0 micrometer). A higher-reflectance, relatively red material occurs as a distinct class of smooth plains that were likely emplaced volcanically; a lower-reflectance material with a lesser spectral slope may represent a distinct crustal component enriched in opaque minerals, possibly more common at depth. A spectrally intermediate terrain probably forms most of the upper crust. Three other spectrally distinct but spatially restricted units include fresh crater ejecta less affected by space weathering than other surface materials; high-reflectance deposits seen in some crater floors; and moderately high-reflectance, relatively reddish material associated with rimless depressions.

The MESSENGER spacecraft encountered Mercury on 14 January 2008, and the Mercury Dual Imaging System (MDIS) (*1*) acquired monochrome, 0.75-μm narrow-angle camera (NAC) and 11-color, 0.4- to 1.0-μm wide-angle camera (WAC) images of parts of Mercury never before seen by a spacecraft. We used WAC color observations (~5 km/pixel) in conjunction with NAC high-resolution (~200 to 500 m/pixel) monochrome images to explore the dominant sources of spectral heterogeneity on Mercury's surface and their correlation with morphologic features.

Early ground-based observations showed that Mercury had a low disk-integrated reflectance and a relatively featureless, positively sloping spectrum across the visible to near-infrared (near-IR) wavelengths (*2*). The lack of resolvable near-IR absorptions led to the conclusion that Mercury's surface is low in ferrous iron (*3*–*5*). Reflectance and mid-IR emission spectroscopy led to the hypothesis that Mercury's upper crust is predominantly composed of plagioclase feldspar, likely with lesser amounts of low-iron pyroxene (enstatite and diopside) and olivine (forsterite) (*4*, *6*, *7*). The surface of Mercury is expected to be heavily altered through space weathering processes that suppress absorption features, lower reflectance, and increase spectral slope (redden the spectrum), thus complicating interpretations of spectral data (*8*). Optically immature materials that are excavated as ejecta and rays from young craters are typically found in deposits less than 100 km in diameter. These smaller deposits contain the less-altered spectral signature of underlying rock and thus provide more definitive information about the composition of the crust and the nature of space weathering processes. The Mariner 10 spacecraft provided the first high-resolution panchromatic and color observations of Mercury (*3*). Similarly to the Moon, crater rays exhibited higher reflectance and shallower slope at visible wavelengths, indicating that space weathering processes operated on Mercury (*3*, *9*, *10*). A major surprise from the Mariner 10 multispectral observations was the lack of reflectance contrast between Mercury's smooth plains, which resemble lunar maria, and the adjacent terrain

[1]School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA. [2]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [3]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [4]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA. [5]National Museum of Natural History, Smithsonian Institution, Washington, DC 20560–0119, USA. [6]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC, 20015, USA. [7]National Air and Space Museum, Smithsonian Institution, Washington, DC 20013, USA.

*To whom correspondence should be addressed. E-mail: robinson@ser.asu.edu

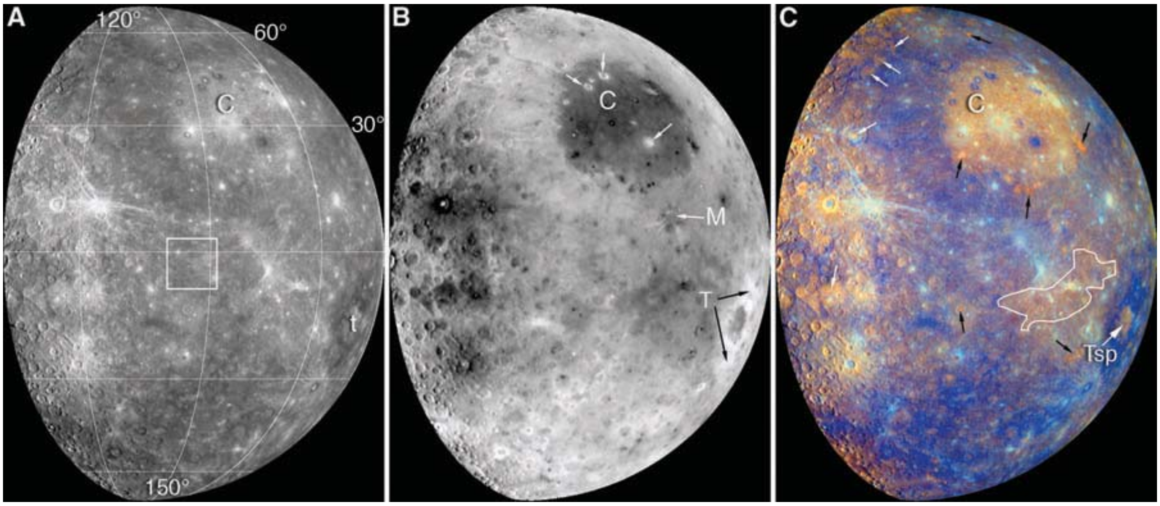


Fig. 1. MDIS WAC departure color sequence (EW0108829678 - EW0108829728C). (**A**) Photometrically corrected 560-nm-filter image. Largest reflectance contrasts are related to immature crater materials, the Caloris smooth plains (C), and Tolstoj basin (t) LRM. White box indicates standard spectral area used to normalize extracted spectra (Fig. 3A). (**B**) PC2 delimits regions with greatest color differences. Caloris smooth plains (C) and ejecta associated with Tolstoj basin (T) exhibit the greatest contrasts not associated with maturity variations. Low-reflectance craters (white arrows) in Caloris basin have the same high values in the PC2 image as Tolstoj ejecta (black arrows). (**C**) Color composite of spectral parameters used to separate units. Red is the inverse of PC2, green is PC1, and blue is relative visible color (430-nm/560-nm ratio). Unlabeled white arrows and irregular polygon indicate relatively young smooth plains deposits that exhibit clear spectral boundaries with basement materials (similar to Caloris plains, C), here interpreted to be volcanic in origin. Tsp denotes HRP within Tolstoj. Black arrows indicate red units interpreted as small volcanic centers.

(*3*, *9*, *10*). Initial analysis of color ratio images led to the conclusion that, unlike on the Moon, color units on Mercury are not well correlated with geomorphic units (*3*, *9*, *10*). Later work based on an improved radiometric calibration of the Mariner 10 color images indicated that at least a subset of smooth plains on Mercury do indeed correspond to color unit boundaries, which bolsters the hypothesis that some of the smooth plains deposits were emplaced as volcanic flows (*11*, *12*).

The MDIS image data (Figs. 1 and 2) were calibrated to radiance factor [known as reflectance, or I/F, observed radiance divided by solar irradiance from a normally solar-illuminated Lambertian disk (*13*)], map-projected, and photometrically corrected to standard viewing geometry (30° solar incidence and 0° emission angles) using a Hapke function with parameters (*14*) derived from Earth-based telescopic measurements. Principal component (PC) analysis was used to identify variations in the 11-band multispectral data set and to identify regions with like spectral properties. In this case, the first principal component (PC1) predominantly maps reflectance variations, whereas the second (PC2) discriminates color variations. Discrete reflectance boundaries in the PC2 map primarily correspond to morphologic boundaries and thus indicate variations in physical state or chemistry (Fig. 1B). To isolate color differences further, we computed color ratios and compared them with the PC maps. We determined distinct regions from these parameter maps (Fig. 1) and extracted representative spectra for analysis (Fig. 3).

MDIS revealed previously unimaged terrain whose reflectance and color properties are similar to those of the region imaged by Mariner 10 (*10*–*12*). The three areally dominant spectral units are low-reflectance material (LRM), moderate-to-high-reflectance smooth plains (HRP), and spectrally intermediate terrain (IT). The three units differ primarily in reflectance and share a red, lunar-like spectral slope indicative of mature, space-weathered soils. LRM (as much as ~30% lower than average reflectance) is widespread (Figs. 1 and 2); the most conspicuous exposures are found in the Tolstoj basin region (Fig. 2D), on some craters—most notably within the Caloris basin (Fig. 1, Fig. 2B)—and in the southern, heavily cratered parts of the newly imaged region (Fig. 1). In some instances, impact events excavated LRM from depth, and this material can be traced outward from the crater in the ejecta blanket (Fig. 2, C, E, and F). In all instances, the LRM has a moderate to shallow spectral slope (Fig. 3) and a shallow minimum near 600 nm. A broad annulus of somewhat dark material exterior to the Caloris basin is ~10% lower in reflectance than the planetary average, with complex interfingered IT and LRM occurrences. Several large craters within the Caloris basin are composed of LRM, indicating that a substantial portion of the material beneath the basin interior smooth plains is also LRM. This annulus of mixed material is approximately equivalent to previously mapped basin ejecta facies (*15*). However, crater counts for a portion of the annulus show that its age is less than that of the Caloris interior plains (*16*). HRP, exemplified by those plains that cover the floor of the Caloris basin, typically have reflectances ~10% above the hemispheric average and exhibit a slightly steeper (redder) spectral slope. Most of the heavily cratered terrain has reflectance properties intermediate to those between those of HRP and LRM.

Three other spectrally distinct but spatially restricted units are also recognized: fresh crater ejecta, bright crater-floor deposits (BCFDs), and moderately high-reflectance relatively reddish material (red spots) associated with rimless depressions. The immature crater ejecta are materials with reflectance elevated as much as 70% above IT and are associated with the walls, floors, and rays of young impact craters (Fig. 2A). These higher-reflectance materials also have a less-red spectral slope (Fig. 3), as do fresh lunar crater materials (*9*, *10*), suggesting that their optical properties have been less affected by space weathering. The BCFDs occur in distinct patches with abrupt boundaries and are found within only a few craters (Fig. 2B). Unlike fresh crater ejecta, the BCFD units do not appear to be associated with exposure of fresh material from the impact that formed the host crater. Similar bright floor deposits were identified in Mariner 10 images (*17*, *18*). This same material is also found in massifs of the peak rings of Raditladi and Eminescu basins. The peak-ring massifs are surrounded by aureoles of comparable high-reflectance material that extend radially for several kilometers. The red spots have elevated reflectances (up to 30% above average), occurring as diffuse halos surrounding scalloped-edged, rimless depressions along the margin of Caloris basin (*19*, *20*) and in the interiors of several craters of the southern heavily cratered terrain. These latter materials are distinguished from fresh-crater and BCFD materials by their spectral slope, which is the reddest observed by MESSENGER. Mariner 10 also observed materials with similar spectral characteristics, the most conspicuous found in the crater Lermontov (*10*, *11*, *18*), but central depressions were seen in only one (*10*).

A consistent attribute of spectra of all of the units is a lack of evidence for an absorption near 1-μm due to ferrous iron in silicates (Fig. 3A). WAC imaging has a spatial resolution two orders of magnitude higher than ground-based spectra, and it resolves fresh impact materials. To under-

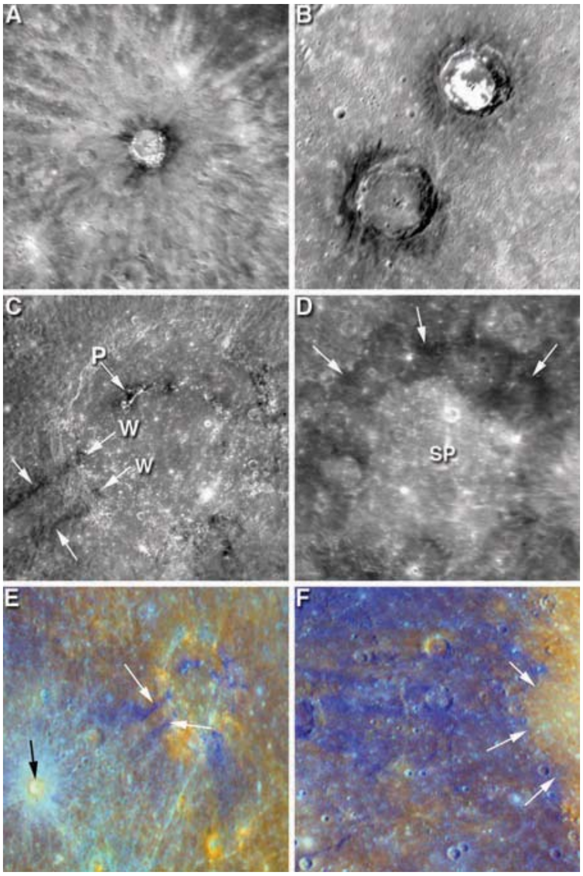


Fig. 2. (**A**) Basho crater (64-km diameter). High-reflectance ray material likely represents immature material, whereas the LRM near the rim may represent immature material of a different composition (EN0108828233M). (**B**) Two impact craters with low-reflectance walls found within the Caloris basin. Sander crater (upper right, 50-km diameter) exhibits BCFDs (EN0108826687M). (**C**) Linear streamers (unlabeled arrows) at Mozart crater (215-km diameter) trace LRM excavated during the impact. Exposures of the same material are visible in the crater wall (W arrows) and in the central peak ring (P) (EN0108827022M). (**D**) Tolstoj HRP (SP) superposed on LRM (arrows); image 825 km wide (EN0108828337M, EN0108828342M). (**E**) Color parameter image highlighting complex color relations. White arrows indicate color anomaly associated with LRM streamers at Mozart crater. Rayed crater at lower left illustrates common occurrence of redder material (black arrow) on the rim of immature craters; image ~800 km wide. (**F**) Western boundary of Caloris HRP (arrows) and basin rim. Note the radial color texture in the lower reflectance annular plains; image ~600 km wide [frame numbers and color scheme for (E) and (F) in Fig. 1 caption]. North is toward the top in all images.

stand the origin of Mercury's reflectance variations despite the low abundance of ferrous iron in silicate, it is informative to compare Mercury's reflectance to that of the Moon. Mercury's geometric reflectance is reported to range from 10% higher to 15% lower than the Moon (*14*), and recent estimates over a broad range of phase angles indicate that the Moon's reflectance is 7 to 17% higher than Mercury's, depending on phase angle (*21*). MESSENGER acquired low-resolution NAC (25 km/pixel) images of the Moon that provide a valuable reference for comparing with NAC Mercury images because the phase angle was nearly identical (65.8° for Mercury, 65.1° for the Moon), eliminating the need for photometric corrections. Comparison of two small areas (highlands on the Moon at 63.2°S, 275.8°E, and Mercury at 1.6°S, 116.1°E) at nearly identical photometric angles shows that Mercury's reflectance (mature IT) is ~27% lower than the mature lunar highlands area, with ~5 weight percent (wt %) FeO (*22*). Previous Moon-Mercury reflectance comparisons were tied to the lunar nearside. To perform a similar comparison, we normalized the Clementine 0.75-µm reflectance map of the Moon to 65° phase angle to match the I/F value at the standard location above, and we determined a lunar nearside average I/F of 0.022 (65° phase). This normalization depends only on the MDIS I/F values, removing uncertainties from the Clementine absolute radiometric calibration from our analysis. A similar approach was used to scale a photometrically normalized NAC 0.75-µm mosaic of the departing hemisphere of Mercury to 66° phase angle, revealing an average I/F of 0.019, a value 14 ± 3% lower than the lunar nearside average. These results are consistent with Earth-based estimates of I/F for the Moon and Mercury that show Mercury's I/F is 8% lower than that of the lunar nearside at 65° phase angle (V band 0.55 µm) (*21*).

Approximately one-third of the lunar nearside is covered by mare basalt containing 16 to 20 wt % FeO and 0 to 13 wt % TiO_2, compared with <6 wt % FeO and <1 wt % TiO_2 for the lunar highlands (*23*). In the absence of substantial iron in silicates, alternatives for lowering reflectance on Mercury include spectrally neutral opaque minerals (*10*) and meteoritic material (*24*). Both can provide a source of ferrous iron for space-weathering-produced submicroscopic metallic iron (known as SMFe or nanophase iron) blebs and coatings (*8*, *24*, *25*). Meteorite impacts are randomly distributed, so if meteoritic material were the source of ferrous iron, the surface of Mercury would be nearly uniform in reflectance except for rays of immature material. The major spectral trend in the data (Fig. 3) is caused by changes in slopes of both the visible and IR parts of the spectrum for fresh craters, consistent with exposure of immature regolith. However, a secondary trend in the data corresponds to a spectral change distinct from regolith maturity, from higher-reflectance smooth plains to LRM (Fig. 3B). This trend is especially evident at the Tolstoj basin. Tolstoj's interior smooth plains are relatively high in reflectance, whereas the ejecta deposits are relatively low in reflectance and less red. We interpret this distinct spectral trend to indicate that a low-reflectance, relatively spectrally neutral component occurs in the upper crust and is admixed in basin ejecta but absent in the smooth plains.

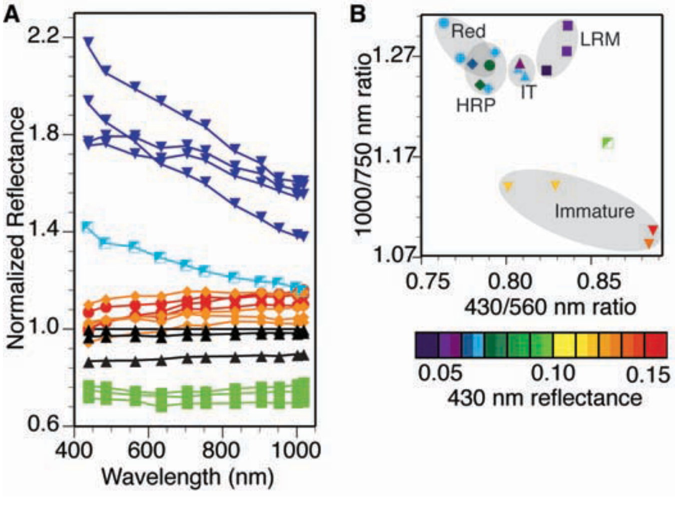


Fig. 3. (**A**) Spectral reflectance of key units discussed in the text, normalized to the standard area shown in Fig. 2A. Blue lines (downward-pointing triangles) are immature material, orange lines (diamonds) are smooth plains (HRP), red lines (circles) are red spots, solid black lines (upward-pointing triangles) are IT, lowest black line is the Caloris annular deposit, green lines (squares) are LRM, and the blue-green (half-filled squares) line is the BCFD found in Sander crater. (**B**) Key ratios used to distinguish spectral units; the single half-filled square corresponds to the Sander BCFD. The lower reflectance IT (diamond) corresponds to the Caloris low-reflectance annulus. The color scale codes the plotted points for absolute reflectance at 430 nm. The 1000-nm/750-nm ratio is a measure of spectral slope in the near-IR. Higher values indicate a steeper (redder) slope. The 430-nm/560-nm ratio is controlled by visible color. A higher value corresponds to a shallower visible slope; an upturn below ~500 nm also contributes to a higher value of this ratio. Generally, mature materials have steep slopes and hence plot in the upper left of the diagram, with immature materials (less steep) found in the lower right. Independent of the maturity trend, variations between major color units can be explained by a variable content of opaque material, which is dark and reduces the spectral slope.

Supporting evidence for a low-reflectance component in portions of Mercury's crust comes from comparison of reflectances of fresh impact materials on Mercury and the Moon. If Mercury's bulk crust is predominantly composed of silicate minerals with low iron and titanium abundances (*4*, *6*, *7*, *9*, *10*), immature material should have relatively high reflectance, perhaps similar to or greater than immature materials found in the low-iron lunar farside highlands (*18*, *26*). Analysis of Mariner 10 images indicated that immature material on Mercury is up to 30% lower in reflectance than comparable immature lunar highland material (*18*). We find similar relations when comparing MDIS NAC mosaics of Mercury and Clementine 0.75-µm mosaics of the Moon, corrected to comparable phase angles as described above and resampled to 5 km/pixel. High-reflectance crater ejecta falls in the reflectance range 0.032 to 0.045 on Mercury, and 0.038 to 0.048 in the lunar highlands. Thus, the reflectance of immature material on Mercury is typically 10% to 20% lower than in the lunar highlands, indicating that Mercury's crust harbors a larger fraction of a low-reflectance component.

What is the nature of Mercury's low-reflectance component? The surface expression of LRM (Figs. 1A and 2) demands endogenic variations within the crust and is not simply an effect of space weathering. Low-reflectance halos interpreted as dark impact melt are observed at lunar craters such as Tycho (*27*), but there is no evidence that the production of impact glass alone could result in the reflectance contrast on the Moon or Mercury, and it is unlikely that the overall reflectance of a major regional unit such as the LRM can be attributed to impact melt. The spectral characteristics of the LRM (shallow visible to near-IR spectral slope and a shallow minimum near 600 nm) (Fig. 3B) are indicative of its composition. Opaque minerals such as ilmenite ($FeTiO_3$) and ülvospinel (Fe_2TiO_4) (*28*, *29*) are low in reflectance, have a flat spectral slope (spectrally neutral), and exhibit a minimum near 600 nm. Such minerals contain ferrous iron, but the 1-µm band is displayed as a broad, weak maximum. Mercury's low reflectance and overall red spectral slope requires that iron be present in some form (either ferrous or metallic or both). The presence of near-ultraviolet heterogeneity and the lack of an obvious 1-µm feature indicative of FeO in silicates are corroborated by the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) surface reflectance spectra (*30*). The relatively flat spectral slope and the shallow minimum in normalized reflectance near 600 nm for the LRM are also seen in the MASCS data (*30*). From these observations and previous work (*2–7*, *10*, *18*, *21*), we propose that Mercury's crust is composed of iron-poor calcium-magnesium silicates (e.g., plagioclase, enstatite, pigeonite, and diopside) with a detectable component of spectrally neutral opaque minerals. Ilmenite ($FeTiO_3$) is the most likely candidate based on lunar analogy and cosmochemical abundance considerations. This lithology is broadly basaltic to gabbroic, with nearly all ferrous iron contained in opaque minerals and not in silicates, requiring redox conditions at or near the iron-wüstite buffer. An alternative to ilmenite is native iron metal, but such a material is relatively red at typical lunar regolith grain sizes (*29*, *31*) and implies more reducing conditions than iron-wüstite. The distinctively red smooth plains (HRP) appear to be large-scale volcanic deposits stratigraphically equivalent to the lunar maria (*20*), and their spectral properties (steeper spectral slope) are consistent with magma depleted in opaque materials. The large areal extent ($>10^6$ km^2) of the Caloris HRP is inconsistent with the hypothesis that volcanism was probably shallow and local (*10*); rather, such volcanism was likely a product of extensive partial melting of the upper mantle.

References and Notes

1. S. E. Hawkins III *et al.*, *Space Sci. Rev.* **131**, 247 (2007).
2. T. B. McCord, J. B. Adams, *Icarus* **17**, 585 (1972).
3. B. Hapke, G. E. Danielson Jr., K. Klaasen, L. Wilson, *J. Geophys. Res.* **80**, 2431 (1975).
4. J. Warell, *Icarus* **161**, 199 (2003).
5. F. Vilas, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Mathews, Eds. (Univ. of Arizona Press, Tucson, 1988), pp. 59–76.
6. A. L. Sprague, T. L. Roush, *Icarus* **133**, 174 (1998).
7. J. Warell, D. T. Blewett, *Icarus* **168**, 257 (2004).
8. B. Hapke, *J. Geophys. Res.* **106**, 10039 (2001).
9. B. Hapke, C. Christman, B. Rava, J. Mosher, *Proc. Lunar Planet Sci. Conf.* **11**, 817 (1980).
10. B. Rava, B. Hapke, *Icarus* **71**, 397 (1987).
11. M. S. Robinson, P. G. Lucey, *Science* **275**, 197 (1997).
12. M. S. Robinson, G. J. Taylor, *Meteorit. Planet. Sci.* **36**, 841 (2001).
13. M. Minnaert, in *Planets and Satellites*, G. P. Kuiper, B. M. Middlehurst, Eds. (Univ. of Chicago Press, Chicago, 1961), pp. 213–248.
14. J. Veverka, P. Helfenstein, B. Hapke, J. D. Goguen, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, 1988), pp. 37–58.
15. P. D. Spudis, J. E. Guest, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Mathews, Eds. (Univ. Arizona Press, Tucson, 1988), pp. 118–164.
16. R. G. Strom *et al.*, *Science* **321**, 79 (2008).
17. D. Dzurisin, *Geophys. Res. Lett.* **4**, 383 (1977).
18. B. W. Denevi, M. S. Robinson, *Icarus*, in press 10.1016/j.icarus.2008.04.021 (2008).
19. J. W. Head *et al.*, *Science* **321**, 69 (2008).
20. S. L. Murchie *et al.*, *Science* **321**, 73 (2008).
21. J. Warell, *Icarus* **167**, 271 (2004).
22. P. G. Lucey, D. T. Blewett, B. L. Jolliff, *J. Geophys. Res.* **105**, 20297 (2000).
23. G. Heiken, D. T. Vaniman, B. M. French, *Lunar Sourcebook: A User's Guide to the Moon* (Cambridge Univ. Press, New York, 1991).
24. S. K. Noble, C. M. Pieters, *Sol. Syst. Res.* **37**, 31 (2003).
25. M. J. Cintala, *J. Geophys. Res.* **97**, 947 (1992).
26. D. T. Blewett, B. R. Hawke, P. G. Lucey, *Meteorit. Planet. Sci.* **37**, 1245 (2002).
27. B. R. Hawke, P. G. Lucey, J. F. Bell, R. Jaumann, G. Neukum, *Lunar Planet. Sci.* **17**, 999 (1986).
28. M. A. Riner, M. S. Robinson, J. A. Tangeman, R. C. Elphic, *Lunar Planet. Sci*. **36**, abstract 1943 (2005).
29. E. A. Cloutis et al., *Icarus*, in press; doi:10.1016/j.icarus.2008.04.018 (2008).
30. W. E. McClintock *et al.*, *Science* **321**, 62 (2008).
31. E. A. Cloutis, D. T. Bailey, M. A. Craig, P. S. Harderson, *Lunar Planet. Sci*. **39**, abstract 1082 (2008).
32. The hundreds of engineers and technical support personnel who brought MESSENGER from a concept to a successful flight project warrant the sustained appreciation of the mission science team. N. Laslo, H. Kang, R. Vaughan, A. Harch, R. Shelton, and A. Berman designed the imaging sequences that made this contribution possible. B. Denevi, K. Becker, and C. Hash are gratefully acknowledged for data calibration and processing. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington.

6 May 2008; accepted 3 June 2008
10.1126/science.1160080

REPORT

Volcanism on Mercury: Evidence from the First MESSENGER Flyby

James W. Head,[1]* Scott L. Murchie,[2] Louise M. Prockter,[2] Mark S. Robinson,[3] Sean C. Solomon,[4] Robert G. Strom,[5] Clark R. Chapman,[6] Thomas R. Watters,[7] William E. McClintock,[8] David T. Blewett,[2] Jeffrey J. Gillis-Davis[9]

The origin of plains on Mercury, whether by volcanic flooding or impact ejecta ponding, has been controversial since the Mariner 10 flybys (1974–75). High-resolution images (down to 150 meters per pixel) obtained during the first MESSENGER flyby show evidence for volcanic vents around the Caloris basin inner margin and demonstrate that plains were emplaced sequentially inside and adjacent to numerous large impact craters, to thicknesses in excess of several kilometers. Radial graben and a floor-fractured crater may indicate intrusive activity. These observations, coupled with additional evidence from color images and impact crater size-frequency distributions, support a volcanic origin for several regions of plains and substantiate the important role of volcanism in the geological history of Mercury.

Volcanic deposits provide important clues to mantle composition, the location of past interior thermal anomalies, and the general thermal evolution of a planet. Relative to the other terrestrial planets, little is known with certainty concerning the history of volcanism on Mercury. In the more than three decades since Mariner 10 flew by Mercury in 1974–75, debate has persisted about the presence or absence of volcanic deposits (*1*), a question we address with data from the first flyby of Mercury by MESSENGER.

Whereas the Moon has distinctive composition-related differences in reflectance between highlands and volcanic maria, the reflectance of Mercury and of its smooth plains is relatively uniform, and therefore the role of volcanism is less obvious. Mariner 10 images, obtained shortly after the Apollo 16 mission to the Moon, revealed two widespread plains units that are similar to the lunar Cayley light plains (*2*, *3*): smooth plains and intercrater plains (*4*). These plains deposits on Mercury were interpreted to be volcanic in origin, on the basis of their smoothness and apparent ponding and embayment of lowland terrain (*4*, *5*). Other researchers, influenced by Apollo 16 results showing that similar-appearing deposits on the Moon were impact breccias, argued that Mercury's smooth plains represented basin ejecta deposits (*3*, *6*). The relatively low resolution of Mariner 10 images (typically ~1 km per pixel) was insufficient to resolve this issue. Lunar-like volcanic features, such as small shields, cones, sinuous rilles, and other vent-related structures, were not detected in Mariner 10 images (*7*, *8*), and the features that were observed did not provide conclusive evidence of a volcanic origin. For example, lobate fronts exposed at the edge of smooth plains suggested the presence of volcanic flow margins on Mercury, but similar features have been identified on the margins of lunar basin ejecta flows (*8*, *9*). The density of impact craters on Caloris

[1]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [2]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [3]School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA. [4]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA. [5]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [6]Southwest Research Institute, 1050 Walnut Street, Boulder, CO 80302, USA. [7]Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Washington, DC 20560, USA. [8]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA. [9]Hawaii Institute of Geophysics and Planetology, University of Hawaii, Honolulu, HI 96822, USA.

*To whom correspondence should be addressed. E-mail: James_Head@brown.edu

basin ejecta and nearby smooth plains deposits suggested that the plains were emplaced after the Caloris basin had formed (*10*), providing evidence that they might be volcanic in origin, not contemporaneous ejecta emplacement. Some lunar Cayley plains, however, display younger ages than adjacent basin ejecta (*11*). Reprocessed Mariner 10 color data (*12*) show color boundaries for smooth plains units, suggesting a distinctive mineralogy. Theoretical studies (*8*, *9*) indicate that a thick low-density crust could inhibit effusive eruptions, particularly if aided by global compression (*13*). Thus, conclusive evidence was lacking for the origin of smooth and intercrater plains and for establishing the role of volcanism in the history of Mercury (*1*, *14*).

The first MESSENGER flyby of Mercury has provided high-resolution multispectral images of regions both seen and unseen by Mariner 10, including the interior plains of the Caloris basin, the circum-Caloris region characterized by smooth plains, and regions of smooth and intercrater plains near the terminator. The data show a variety of evidence for volcanic features and deposits across this part of the planet. The new images (Figs. 1 to 4) reveal several volcanic vents and related deposits distributed along the southern margins of the Caloris basin interior and characterized by distinctive color anomalies (*15*, *16*) (Fig. 1). Several irregularly shaped rimless depressions are surrounded by smoother deposits that have albedos differing from those of surrounding units. For example, along the southern margin of Caloris, a kidney-shaped depression about 20 km long is centered on a smooth deposit that is greater than 100 km in diameter (Fig. 1). A relatively bright deposit that surrounds the kidney-shaped depression to a radial distance of about 25 km appears to discolor or mantle the underlying terrain and becomes more diffuse toward its edges. Several small irregular depressions are observed in the southwestern part of the smooth deposit. Each has an irregular shape similar to those of volcanic vents on other planetary bodies, unlike the generally circular shapes of primary craters and elongated secondary craters with rims. These features and their close association with the distinctive smooth deposits imply that these broad structures are volcanic vents.

Although altimetry data are not available for this region, several lines of evidence suggest that the broad feature is domelike in nature (Fig. 1). The decrease in brightness along the northwestern deposit is consistent with a surface sloping away from the kidney-shaped depression in the downsun direction. Small massifs at the Caloris basin rim to the southwest appear to be partially buried, which suggests that they are covered by volcanic deposits from this region. Finally, two large impact craters (~25 to 30 km across) straddle the northwestern edge of the deposit. Both contain interior deposits (walls and rough floors) that are typical of impact craters of this size (*17*) and seen elsewhere in Fig. 1. The rim and exterior deposits differ between the two craters, however: The eastern rim of the southernmost crater is heavily embayed (toward the central part of the feature) such that its rim crest is no longer visible. Rim heights of fresh Mercury craters of the same size (*17*) imply that about 800 m of lava would be required to flood the crater to the level of the rim crest. Thus, on the basis of these lines of evidence, we interpret this structure to be a broad, low shield volcano similar in character to mare domes on the Moon that erupted a sequence of lava flows (*18*). At least four similar examples are seen in the southern part of the interior Caloris basin rim (*15*). The diffuse-bordered bright halo surrounding the kidney-shaped depression in Fig. 1 is interpreted to consist of pyroclastic deposits emplaced from vent-centered eruptions, on the grounds that similar kidney-shaped depressions and pyroclastic deposits have been documented on the Moon (*19*).

MESSENGER images show that many impact craters and areas between craters have been flooded with lava. Large fresh impact craters (Fig. 2A) have rough flat floors; central peaks and peak rings; terraced walls; and—beyond the interior—a sharp rim crest, raised rim, and radially textured ejecta. Crater chains and crater clusters are seen in many exterior ejecta deposits. Volcanically modified impact craters on the Moon, such as Archimedes crater within the Imbrium basin (*20*), have a shallow, smooth crater floor, a lack of peaks and peak rings, and flooded and embayed exterior deposits.

The morphology of a large, degraded ~240-km-diameter impact crater near the terminator offers an opportunity to compare fresh and degraded craters and the processes responsible for their modification (Fig. 2B). The crater is located far from the Caloris basin rim (1375 km to the southwest), and there is no radial texture suggestive of ejecta emplacement from a nearby impact basin. Smooth plains nearly fill the crater and also embay the interior crater walls, which is suggestive of volcanic flooding. Five smaller impact craters are superposed on the degraded crater: Three lie on the rim crest, one lies outside the crater (to the northwest), and one lies on the southern part of the crater floor. In contrast to fresh craters (Fig. 2A), the 85-km-diameter crater on the southern crater floor has been embayed along its northern margin up to the base of the raised rim, and all of its exterior ejecta have been buried by the emplacement of plains, which also flood its interior. Similar relations are seen on the three craters on the degraded crater rim. Crater floor plains embay the exterior ejecta of each of these craters, up to the base of their respective raised rims, although exterior ejecta deposits are preserved elsewhere on the degraded crater rim. Ejecta from the large superposed crater to the northwest, once present on the floor of the degraded crater, are now buried by smooth plains. Small “ghost” craters, apparently buried by plains emplacement, are also observed (Fig. 2C).

Fresh craters with diameters similar to that of the large degraded crater are generally more than 5 km deep (*17*), and yet this crater has undergone substantial shallowing. The degraded state of the crater and the superposition of numerous large craters on it imply an extended interval of modification. Moreover, the embayment and burial of the exterior deposits of each of these craters by smooth plains suggest that more recent and sequential volcanic flooding was involved. In addition, the presence of flowlike scarps and ghost craters, the lack of basin ejecta sculpture, and the distance from known basins all combine to favor a volcanic origin for the emplacement of these

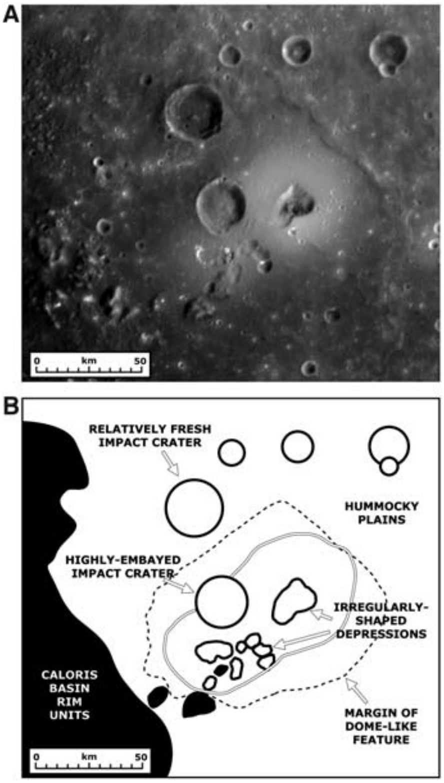

A
0 km 50
B
RELATIVELY FRESH IMPACT CRATER
HUMMOCKY PLAINS
HIGHLY-EMBAYED IMPACT CRATER
IRREGULARLY-SHAPED DEPRESSIONS
CALORIS BASIN RIM UNITS
MARGIN OF DOME-LIKE FEATURE
0 km 50

Fig. 1. Central kidney-shaped depression superposed on a broad, smooth domelike feature and surrounded by bright halo material. As one of several similar features along the inner margin of the southern Caloris basin rim [bottom left in (A) and (B)], this feature resembles small shield volcanoes on the Moon (*18*) and is interpreted as a volcanic vent. (**A**) Dome and kidney-shaped depression [mosaic of Mercury Dual Imaging System narrow-angle camera (MDIS NAC) images EN0108826812M and EN0108826877M, centered at 22.5°N, 146.2°E]. North is to the top in this and subsequent images. (**B**) Sketch map illustrating the main features and relationships. The paired lines enclose bright material surrounding depressions. This and the following images are at 750-nm wavelength. Locations on Mercury of this and the following images are shown in the supporting online material of (*29*).

crater interior plains. Subsequent to flooding, a large northeast-to-southwest–trending fault scarp (Fig. 2, B and C) cut the interior superposed crater and the southern part of the degraded crater.

On the Moon and Mars, volcanically flooded impact craters are often outlined by circular wrinkle-ridge patterns (*21*) (Fig. 3B). The rim crests of impact craters that are almost completely flooded by lavas act to localize deformation of the volcanic sequence, forming the wrinkle-ridge ring. MESSENGER near-terminator image data reveal a large (~60-km-diameter) wrinkle-ridge ring in an area of extensive smooth plains (Fig. 3A). Several small knobs, interpreted to be the remnants of the crater rim crest, are all that remain of the crater. This structure interrupts a regional northeast-trending pattern of wrinkle ridges. If the current ring diameter (60 km) is a good approximation of the original size of the crater, about 2.7 km of lava fill would be required to flood the original crater to its rim crest (*17*). Several other wrinkle-ridge rings are observed in this region. These observations further attest to the emplacement of substantial thicknesses of volcanic plains.

Intrusive activity is a prerequisite for volcanism, and the MESSENGER data reveal possible candidates for shallow sills and dikes. On the Moon, fractures on crater floors have been interpreted to form either by viscous relaxation (*22*) or the intrusion of sills and uplift of crater floors (*23*). The presence of floor-fractured craters was suspected from Mariner 10 data (*24*), and the MESSENGER data confirm their presence (Fig. 4). A ~35-km-diameter crater, located near the margins of the extensive plains units, shows two fractured domes on its floor (Fig. 4). The local-

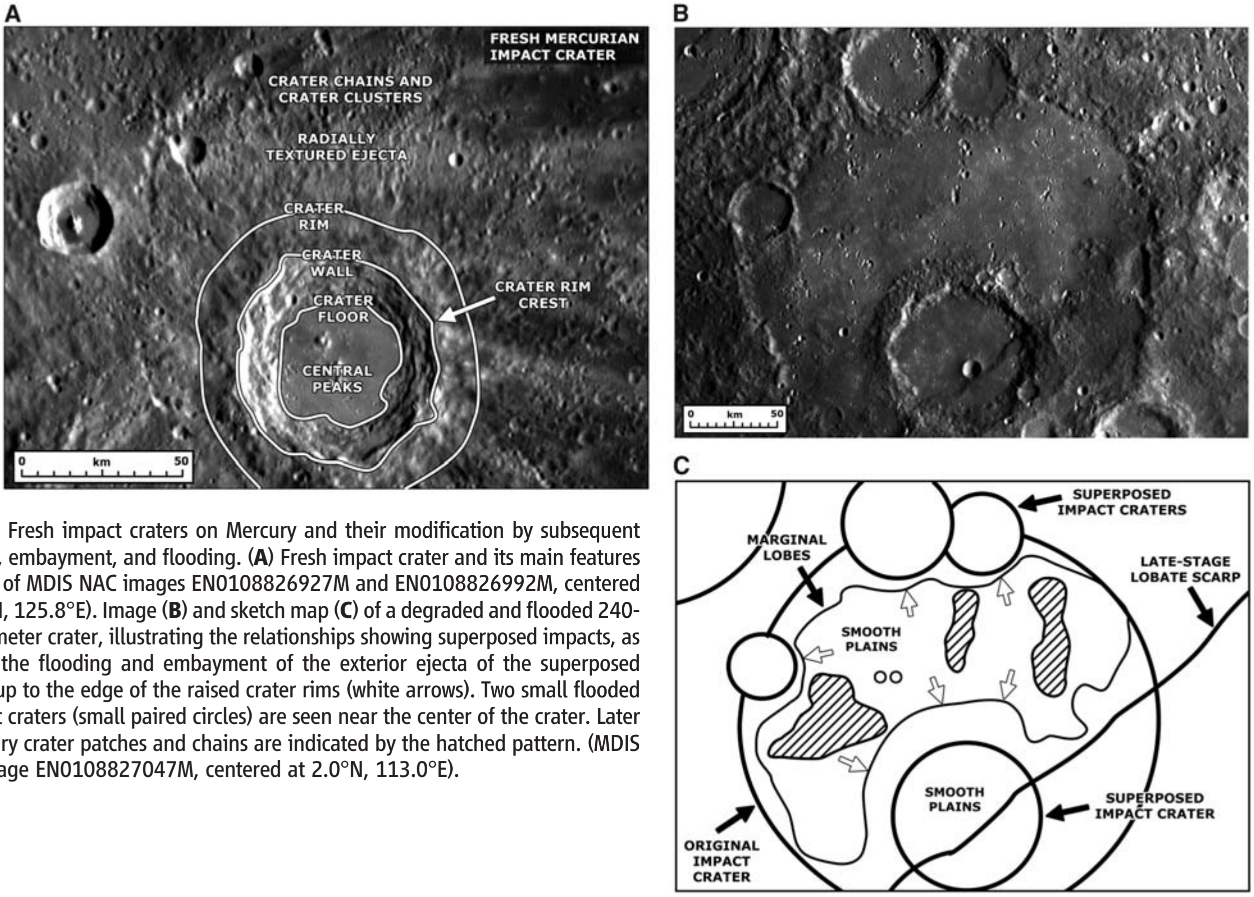


Fig. 2. Fresh impact craters on Mercury and their modification by subsequent impacts, embayment, and flooding. (**A**) Fresh impact crater and its main features (mosaic of MDIS NAC images EN0108826927M and EN0108826992M, centered at 9.6°N, 125.8°E). Image (**B**) and sketch map (**C**) of a degraded and flooded 240-km-diameter crater, illustrating the relationships showing superposed impacts, as well as the flooding and embayment of the exterior ejecta of the superposed craters up to the edge of the raised crater rims (white arrows). Two small flooded or ghost craters (small paired circles) are seen near the center of the crater. Later secondary crater patches and chains are indicated by the hatched pattern. (MDIS NAC image EN0108827047M, centered at 2.0°N, 113.0°E).

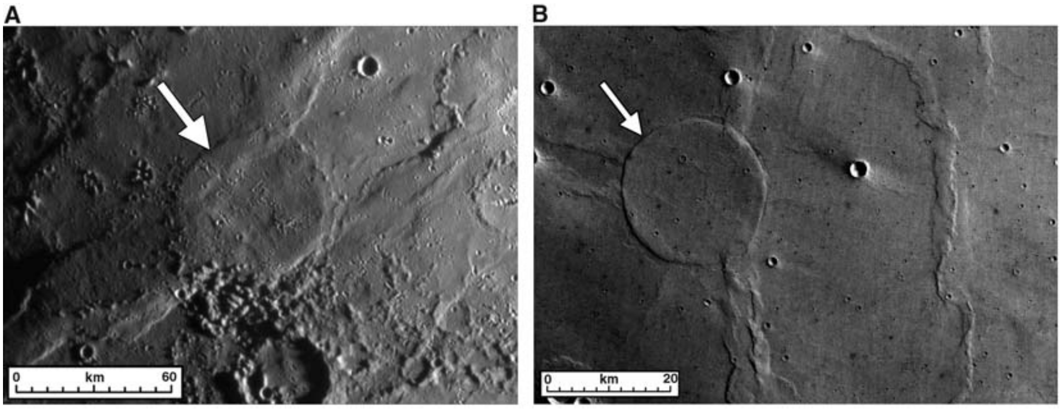


Fig. 3. Wrinkle-ridge rings on Mercury and Mars. These features are interpreted to be impact craters almost completely flooded by lavas; subsequent regional deformation forms linear wrinkle ridges whose regional patterns are disrupted into rings by the presence of the buried crater rim. (**A**) Near-terminator view of a ~60-km-diameter wrinkle-ridge ring in an area of extensive smooth plains (MDIS NAC image EN0108826972M, centered at 10.0°N, 98.4°E). (**B**) A similar wrinkle-ridge ring in volcanic plains in Hesperia Planum, Mars (30°S, 115°E) (portion of Mars Express High-Resolution Stereo Camera nadir image h2660_0001).

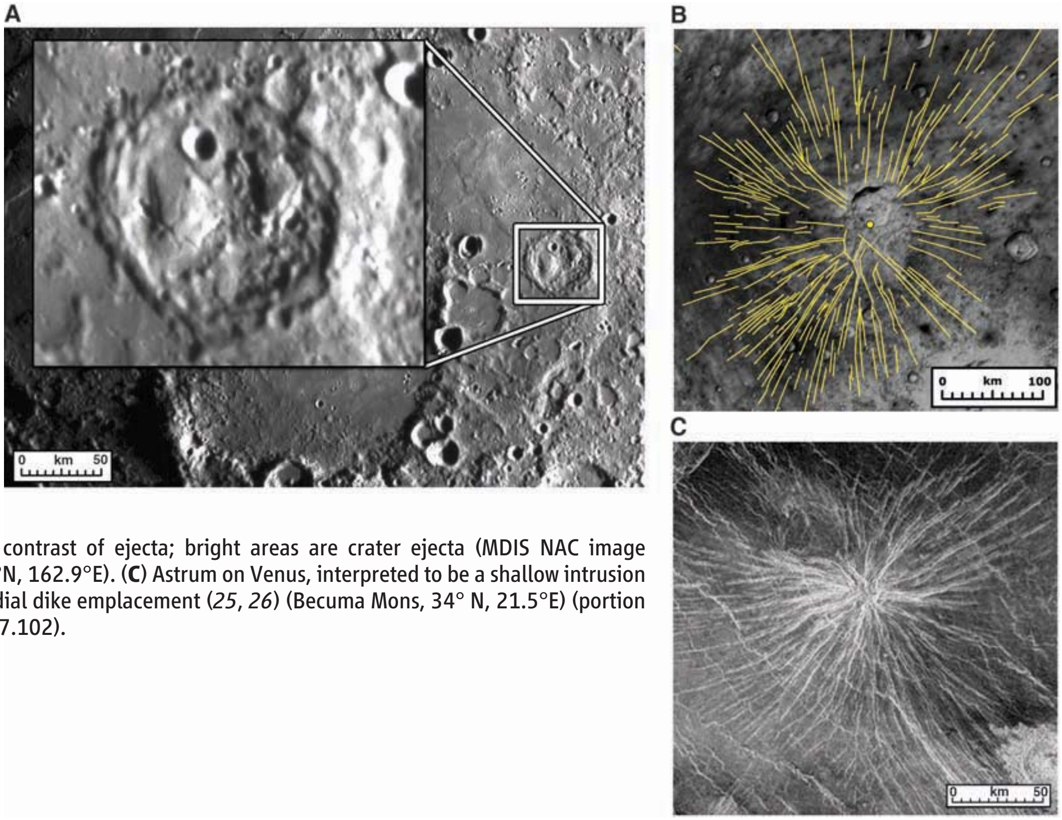


Fig. 4. Floor-fractured crater and radial graben structure, suggestive of intrusive processes on Mercury, as compared with a structure on Venus. (**A**) Floor-fractured, ~35-km-diameter crater (inset); two mounds on the crater floor show networks of fractures. The floor-fractured crater is near extensive tracts of smooth plains to the west and northwest, interpreted to be of extrusive volcanic origin (Fig. 3A) (mosaic of MDIS NAC images EN0108826972M and EN0108826977M, centered at 6.5°N, 100.5°E). Radial graben structure (Pantheon Fossae) on Mercury and astrum on Venus: (**B**) Sketch map of graben radiating from near the center of the Caloris basin, superposed on MESSENGER MDIS image shown as negative to enhance the visual contrast of ejecta; bright areas are crater ejecta (MDIS NAC image EN0108828540M, centered at 29.9°N, 162.9°E). (**C**) Astrum on Venus, interpreted to be a shallow intrusion surrounded by graben formed by radial dike emplacement (*25*, *26*) (Becuma Mons, 34° N, 21.5°E) (portion of Magellan image C1-MIDR.30N027.102).

ized nature of the floor fractures favors intrusion over broader-scale relaxation.

A second candidate for magmatic intrusion is Pantheon Fossae, a radial graben structure in the center of the Caloris basin (*15*). Over 100 graben, typically 2 to 3 km wide and tens to hundreds of kilometers long, radiate away from the center. Several hypotheses are being assessed for the origin of this feature (*15*) (Fig. 4B), but one possible mechanism—which was proposed for morphologically similar features (astra) on Venus (Fig. 4C)—is that rising magma forms a reservoir at a neutral buoyancy zone below the surface, and over-pressurization of the reservoir results in emplacement of radial dikes. Shallow radial dikes imply that the near-surface stress field is extensional; as a result, graben tend to form over the dikes. Under some conditions, dikes will give rise to associated extrusive volcanism, whereas under other conditions, dikes remain largely intrusive (*25*). So far, Pantheon Fossae is the only such radial graben structure observed on Mercury; in contrast, more than 60 astra have been documented on Venus (*26*). The location of the single Mercury example in central Caloris basin might be related to pressure-release melting induced by mantle upwelling localized by basin formation (*27*), in which case additional examples might be sought at the centers of other basins on Mercury.

These data, together with multispectral images and spectra showing evidence for multiple plains units (*16*) of different ages (*28*), substantiate the role of volcanism as an important process in the geological history of Mercury. The volcanic style of Mercury appears to be similar to that of the Moon, with regional volcanic plains filling craters, basins, and intervening areas early in the history of the planet. That style contrasts with the huge volcanic edifices and more extended duration of volcanism on Mars and the plate-boundary and hot spot volcanism on Earth.

References and Notes

1. J. W. Head *et al.*, *Space Sci. Rev.* **131**, 41 (2007).
2. N. J. Trask, J. F. McCauley, *Earth Planet. Sci. Lett.* **14**, 201 (1972).
3. V. R. Oberbeck, *Rev. Geophys. Space Phys.* **13**, 337 (1975).
4. B. C. Murray, R. G. Strom, N. J. Trask, D. E. Gault, *J. Geophys. Res.* **80**, 2508 (1975).
5. R. G. Strom, *Phys. Earth Planet. Inter.* **15**, 156 (1977).
6. D. E. Wilhelms, *Icarus* **28**, 551 (1976).
7. M. C. Malin, *Proc. Lunar Planet. Sci. Conf.* **9**, 3395 (1978).
8. S. M. Milkovich, J. W. Head, L. Wilson, *Meteorit. Planet. Sci.* **37**, 1209 (2002).
9. J. W. Head *et al.*, in *Environmental Effects on Volcanic Eruptions: From Deep Oceans to Deep Space*, T. K. P. Gregg, J. R. Zimbelman, Eds. (Plenum, New York, 2000), pp. 143–178.
10. P. D. Spudis, J. E. Guest, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 118–164.
11. D. E. Wilhelms, *U.S. Geol. Surv. Prof. Pap.* **1348**, 1 (1987).
12. M. S. Robinson, P. G. Lucey, *Science* **275**, 197 (1997).
13. R. G. Strom, N. J. Trask, J. E. Guest, *J. Geophys. Res.* **80**, 2478 (1975).
14. S. C. Solomon, R. L. McNutt Jr., R. E. Gold, D. L. Domingue, *Space Sci. Rev.* **131**, 3 (2007).
15. S. L. Murchie *et al.*, *Science* **321**, 73 (2008).
16. M. S. Robinson *et al.*, *Science* **321**, 66 (2008).
17. R. J. Pike, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 165–273.
18. J. W. Head, A. Gifford, *Moon Planets* **22**, 235 (1980).
19. B. K. Lucchitta, H. H. Schmitt, *Proc. Lunar Sci. Conf.* **5**, 223 (1974).
20. J. W. Head, *Moon Planets* **26**, 61 (1982).
21. T. R. Watters, *J. Geophys. Res.* **98**, 17049 (1993).
22. J. L. Hall, S. C. Solomon, J. W. Head, *J. Geophys. Res.* **86**, 9537 (1981).
23. P. H. Schultz, *Moon* **15**, 241 (1976).
24. P. H. Schultz, *Phys. Earth Planet. Inter.* **15**, 202 (1977).
25. E. A. Parfitt, J. W. Head, *Earth Moon Planets* **61**, 249 (1993).
26. A. Krassilnikov, J. W. Head, *J. Geophys. Res.* **108**, 5108 (2003).
27. L. T. Elkins-Tanton, B. H. Hager, T. L. Grove, *Earth Planet. Sci. Lett.* **222**, 17 (2004).
28. R. G. Strom *et al.*, *Science* **321**, 79 (2008).
29. S. C. Solomon *et al.*, *Science* **321**, 59 (2008).
30. We thank the MESSENGER mission instrument, planning, and operations teams. R. L. McNutt Jr., D. L. Domingue, and S. E. Hawkins III provided leadership in mission planning and instrument design; N. Laslo, H. Kang, R. Vaughan, A. Harch, R. Shelton, and A. Berman designed the imaging sequences; and B. Denevi, K. Becker, and C. Hash provided data calibration and processing. We thank S. Schneider, J. Dickson, C. Fassett, L. Kerber, D. Hurwitz, G. Morgan, S. Schon, L. Ostrach, and N. Chabot for support. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to the Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington.

17 April 2008; accepted 3 June 2008
10.1126/science.1159256

REPORT

Geology of the Caloris Basin, Mercury: A View from MESSENGER

Scott L. Murchie,[1] Thomas R. Watters,[2] Mark S. Robinson,[3] James W. Head,[4] Robert G. Strom,[5] Clark R. Chapman,[6] Sean C. Solomon,[7] William E. McClintock,[8] Louise M. Prockter,[1] Deborah L. Domingue,[1] David T. Blewett[1]

The Caloris basin, the youngest known large impact basin on Mercury, is revealed in MESSENGER images to be modified by volcanism and deformation in a manner distinct from that of lunar impact basins. The morphology and spatial distribution of basin materials themselves closely match lunar counterparts. Evidence for a volcanic origin of the basin's interior plains includes embayed craters on the basin floor and diffuse deposits surrounding rimless depressions interpreted to be of pyroclastic origin. Unlike lunar maria, the volcanic plains in Caloris are higher in albedo than surrounding basin materials and lack spectral evidence for ferrous iron-bearing silicates. Tectonic landforms, contractional wrinkle ridges and extensional troughs, have distributions and age relations different from their counterparts in and around lunar basins, indicating a different stress history.

The Caloris basin, the youngest large impact basin known on Mercury, was seen in its entirely during the first encounter of Mercury by the MESSENGER spacecraft in January 2008 (*1*). Caloris provides important information for understanding Mercury's geology because it exposes layering of the planet's crust, and it contains tectonic and volcanic features that are well-preserved as compared with those of older basins more modified by subsequent impact cratering. Imaging coverage of the basin interior and ejecta was therefore a focus of the encounter sequence. Here we make use of color and high-resolution monochrome images to reconstruct the geological evolution of the basin and to assess the origin and distribution of smooth plains material interior to the basin, the compositional stratification of Mercury's upper crust, and the large-scale deformational history of the region.

Over 30 years ago, Mariner 10 imaged the eastern part of Caloris (*2*–*5*), and the basin has subsequently been studied with ground-based radar (*6*). Major units forming the basin include a rim of concentric massifs with intermontane plains (the Caloris Montes and Nervo Formations), which form an annulus varying in width (up to 250 km) surrounding the light-colored basin interior (Fig. 1A). An outlying darker annulus consists of rolling ejecta deposits (the Odin Formation), which grade into radially lineated plains and overlapping secondary craters in clusters, thought to be distal sculpted ejecta (the van Eyck Formation). These formations are comparable to counterparts in lunar basin materials, such as the Montes Rook and Hevelius Formations that surround the Orientale basin (*7*). Two types of smooth plains were recognized in association with Caloris. Exterior to the basin, annular smooth plains east of Caloris exhibit pervasive wrinkle ridges. The interior of the basin contains plains proposed on the basis of Mariner 10 images to have originated either as volcanic flows (*2*, *8*) or impact melt (*9*). In the former case, the interior plains would be equivalent to lavas forming Mare Orientale, and in the latter to the Maunder Formation, which is thought to be impact melt. The interior plains exhibit wrinkle ridges and younger, cross-cutting extensional troughs (*10*–*12*). Wrinkle ridges, thought to have formed by a combination of thrust faulting and folding (*13*–*15*), occur near the eastern basin margin and are both concentric and radial to the basin, a pattern common in mare basalt-filled lunar basins. The troughs are graben formed by extensional stresses and have linear and sinuous segments that form giant polygons (*10*, *11*, *15*). Before MESSENGER, it was not known if or how far the wrinkle ridges and graben extended westward into the then-unimaged portion of the basin, or how consistent their spatial and age relations were across the basin.

Imaging by the Mercury Dual Imaging System (MDIS) (*16*) during MESSENGER's first Mercury flyby was optimized for coverage of Caloris, with a narrow-angle camera mosaic at 200 to 300 m/pixel, and a wide-angle camera 11-color mosaic at 2.4 km/pixel. Both data sets were photometrically corrected to normalized reflectance at a standard geometry (30° incidence angle, 0° emission angle) and map-projected. From the 11-color data, principal component analysis and spectral ratios were used to highlight the most important trends in the data (*17*). The first principal component dominantly represents brightness variations, whereas the second principal component (PC2) isolates the dominant color variation: the slope of the spectral continuum. Several higher components isolate fresh craters; a simple color ratio combines these and highlights all fresh craters. PC2 and a 480-nm/1000-nm color ratio thus represent the major observed spectral variations (Fig. 1B).

The images show that basin exterior materials, including the Caloris Montes, Nervo, and von Eyck Formations, all share similar color properties, which continue with little variation azimuthally around the basin from the part imaged by Mariner 10 (Fig. 1). They have a normalized reflectance of ~0.085 at 560 nm, a red lunarlike spectral continuum, and an absence of light-colored smooth plains that occur as patches in the surrounding highlands (*17*, *18*). Smooth plains within the Odin Formation lack distinctive color properties. In contrast, Caloris interior plains are about 10% higher in normalized reflectance and have a redder spectral continuum. The northwestern half of the interior plains is both slightly redder and slightly higher in albedo than the southeastern half. The interior plains exhibit craters with diameters of several tens of kilometers whose interiors and ejecta have a lower albedo and less red color than the basin exterior and resemble darker terrain exterior to the basin (red circles, Fig. 1A). The northwestern part of the basin interior exhibits craters with similar dark rims but light floors resembling the interior plains (blue circles, Fig. 1A). The margin of the basin contains diffuse, 30- to 100-km-diameter patches of very red material, typically ~40% higher in normalized reflectance than the basin exterior (red patches in Fig. 1B, circled in white). Fresh impact craters with comparably elevated albedo are spectrally distinctive from the red patches, with a less red spectral continuum than other materials (bluer color in Fig. 1B) as is also typical of fresh lunar craters (*19*).

Higher-resolution views of the western annulus (Fig. 1, C to E) show smooth to rolling plains surrounding the edge of the basin, which grade into radially lineated equivalents of the van Eyck Formation, particularly in the northwest (white arrow, Fig. 1C). The northwestern part of the basin interior also exhibits interior plains embaying knobs of the basin rim as well as a partly preserved inner ring, whose eastern portion was suggested in Mariner 10 images but not identified definitively (*2*) (red arrows, Fig. 1C). The western equivalent of the Caloris Montes Formation (red arrows, Fig. 1D) is an inward-facing scarp broken into clusters of knobs, locally higher in normalized

[1] Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [2] Center for Earth and Planetary Studies, National Air and Space Museum, Smithsonian Institution, Washington, DC 20015, USA. [3] Department of Geological Sciences, Arizona State University, Tempe, AZ 85251, USA. [4] Department of Geological Sciences, Brown University, Providence, RI 02906, USA. [5] Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [6] Southwest Research Institute, 1050 Walnut Street, Boulder, CO 80302, USA. [7] Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20650, USA. [8] Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA.

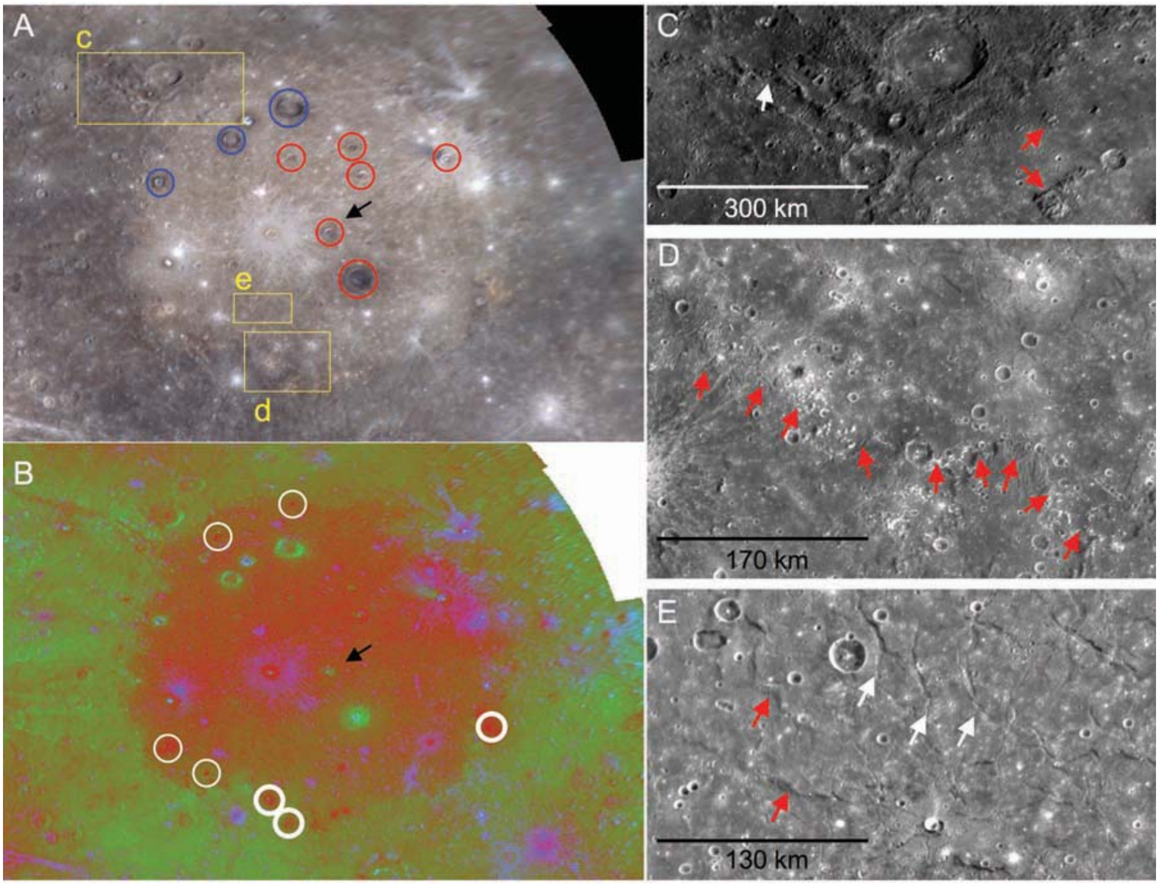


Fig. 1. Overview of the Caloris basin. All images are mosaics in equirectangular projection, and north is up. (**A**) Enhanced (but still subtle) color image map showing 1000-, 750-, and 480-nm images in the red, green, and blue image planes. Yellow boxes are insets shown at higher resolution in (C), (D), and (E); red circles indicate superposed dark craters; blue circles indicate embayed dark craters; and the black arrow indicates the central crater Apollodorus of the radial graben complex. (**B**) Same view after data transformation. PC2, which captures variations between light plains and darker terrain, is shown in the green image plane and inverted in the red plane. The ratio of normalized reflectances at 480 nm/1000 nm, which highlights fresh impact ejecta, is shown in the blue plane. Small red spots are extremely red and elevated in albedo; those in white circles are centered on small rimless depressions. Thicker circles indicate features shown in Fig. 3. (**C**) High-resolution view of inset c in (A), showing radially lineated ejecta (white arrow) and massifs embayed by the interior plains (red arrows). (**D**) High-resolution view of inset d in (A), showing massifs at the inner edge of the western continuation of the Caloris Montes Formation (red arrows). (**E**) High-resolution view of inset e in (A), showing examples of wrinkle ridges (red arrows) and troughs (white arrows) in the southwestern interior plains.

reflectance. Fitting the entirety of Caloris Montes yields a main basin ring diameter of 1550 km, as compared with 1340 km estimated from Mariner 10 data. The western interior plains (Fig. 1E) are deformed by both wrinkle ridges and extensional troughs, similarly to the eastern part of the plains imaged by Mariner 10. The continuity in color and morphology around the basin interior and exterior indicates that the units seen by Mariner 10 extend throughout the rest of the basin and its surroundings.

Two lines of evidence support the idea that the interior plains were emplaced by volcanism, as the stratigraphic equivalent of lunar maria rather than as Caloris impact melt deposits. The first line of evidence is the occurrence within the main basin ring of craters flooded by light-colored interior plains, analogous to lunar basin interior craters such as Archimedes in Mare Imbrium. Some interior craters (red circles, Fig. 1A) have color and morphology consistent with their superposition on the interior plains (left column, Fig. 2): unbroken crater rims and intact central peak rings; a gradual change in ejecta texture with radial distance from the crater rim; and interior color and albedo properties consistent with the crater wall but distinct from surrounding plains, suggesting that an underlying dark layer was excavated. In contrast, craters interpreted as embayed and infilled (blue circles, Fig. 1A, and right column, Fig. 2) have attributes suggesting that they were flooded by interior plains material: breached rims and discontinuous central peak rings, an abrupt change in texture outside the crater where plains lap onto ejecta, and interior color properties distinct from the crater wall but matching the surrounding, lighter plains. Because the formation times of the several large craters on the Caloris floor must have spanned an interval much longer than that required for the emplacement of impact melts, the infilling of these craters must have occurred subsequently. Spudis and Guest (*2*) used Mariner 10 images and similar criteria to interpret a volcanic origin for light-colored interior plains of the Tolstoj basin, located to the southeast of Caloris. Interior flooding and embayment relations similar to these are also seen in smooth plains outside the Caloris basin in the MESSENGER data (*17*, *18*).

The second line of evidence for volcanism is the presence of diffuse bright deposits concentrated along the margin of the basin (Fig. 3), associated with apparent volcanic structures (*18*). The bright deposits have a redder spectral continuum than other materials, in contrast to comparably bright fresh impact craters, which also typically display prominent rays (*19*). Moreover, the bright deposits are centered on irregularly shaped, scallop-rimmed depressions without raised rims. The morphology of these depressions does not resemble that of impact craters but is consistent with volcanic vents observed in the lunar maria (*20*). Similar diffuse deposits on the Moon are pyroclastic in origin (*21*), and we interpret these deposits to be pyroclastic as well. Spectrally similar materials are also observed surrounding somewhat degraded craters on the Caloris interior plains; these craters are smaller than those exposing dark material, suggesting that material resembling the diffuse deposits was excavated from a shallower depth in the interior plains.

Unlike the lunar maria, the Caloris interior plains are higher in albedo than the underlying basin material and lack spectral evidence for

Fig. 2. Examples of morphologic and spectral indicators of craters being superposed on the interior plains (Atget crater, left column) or embayed by interior plains material (unnamed crater, right column). The top row of images is from the narrow-angle camera high-resolution mosaic. The bottom row is transformed wide-angle camera color imagery, using the same representation of the data as in Fig. 1B. All images are mosaics in equirectangular projection, and north is up.

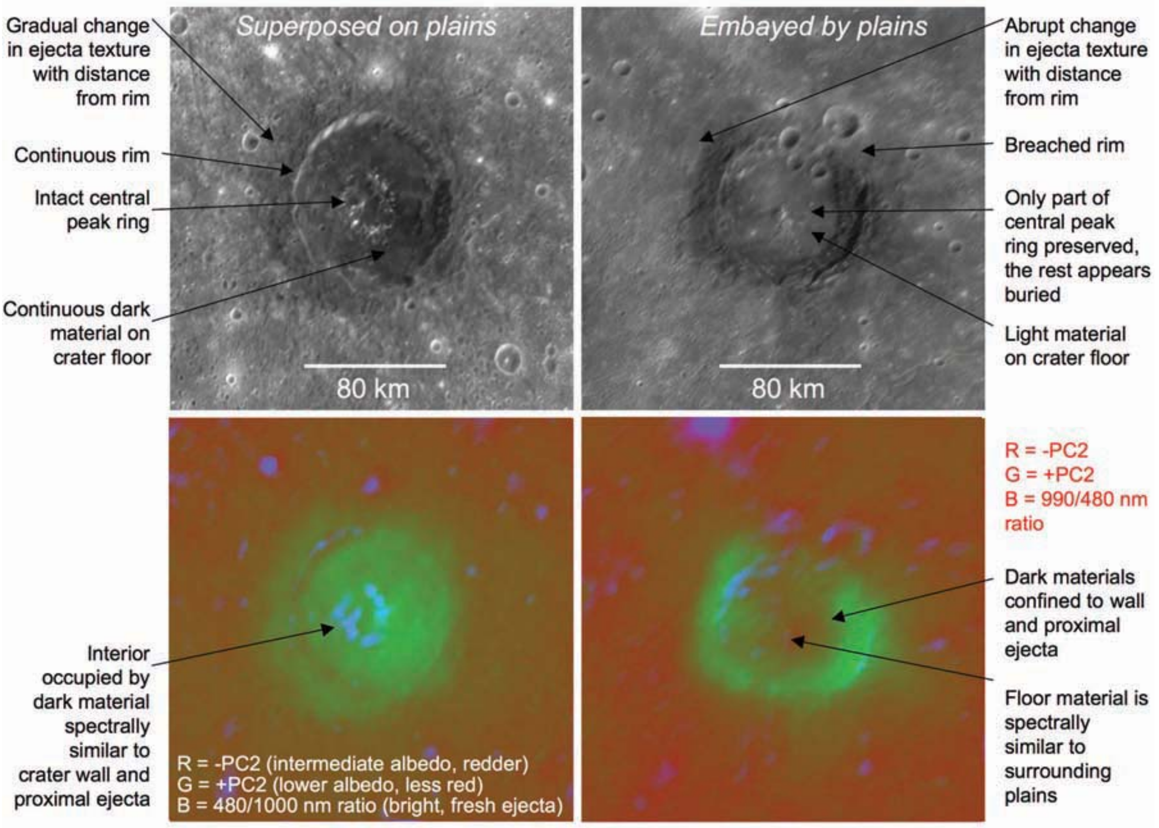
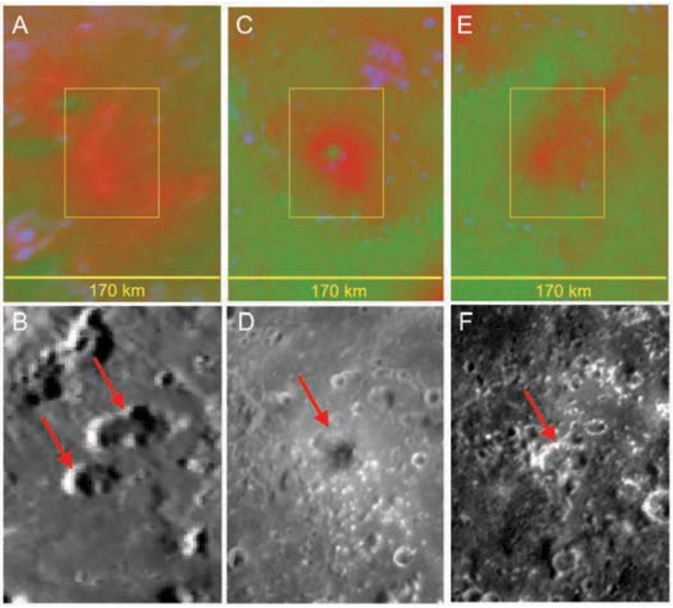

Superposed on plains
Embayed by plains
Gradual change in ejecta texture with distance from rim
Continuous rim
Intact central peak ring
Continuous dark material on crater floor
80 km
80 km
Abrupt change in ejecta texture with distance from rim
Breached rim
Only part of central peak ring preserved, the rest appears buried
Light material on crater floor
R = -PC2
G = +PC2
B = 990/480 nm ratio
Dark materials confined to wall and proximal ejecta
Floor material is spectrally similar to surrounding plains
Interior occupied by dark material spectrally similar to crater wall and proximal ejecta
R = -PC2 (intermediate albedo, redder)
G = +PC2 (lower albedo, less red)
B = 480/1000 nm ratio (bright, fresh ejecta)

Fig. 3. Spectral and morphologic characteristics of the diffuse red spots along the margin of Caloris, for the three spots with thick white outlines in Fig. 1B. The top row (**A**, **C**, and **E**) is transformed wide-angle camera color imagery, using the same representation of the data as in Fig. 1B. Each image in the bottom row (**B**, **D**, and **F**) is of the area shown in the yellow box in the image above. (D) and (F) are from the narrow-angle camera high-resolution mosaic, and (B) is from Mariner 10. Red arrows point to the central rimless, scalloped depressions. All images are mosaics in equirectangular projection, and north is up.



silicates containing ferrous iron (Fe^{2+}). Neither fresh craters, nor the diffuse bright patches, nor any other materials exhibit a resolvable absorption at 0.90 to 1.05 μm (the 1-μm absorption) due to Fe-bearing olivine or pyroxene (*17*). Such an absorption feature is typical of young lunar craters superposed on either the lunar maria or average highlands. The lack of a 1-μm absorption suggests that the content of Fe in silicates is much lower than in the lunar maria (*19*); common igneous silicate minerals consistent with the absence of a 1-μm absorption include Mg-rich olivine, Mg- or Mg-Ca–rich pyroxene, or plagioclase feldspar.

MESSENGER images show that graben orientation progresses from a dominantly radial pattern at the center of the basin to a polygonal pattern at the edge of the basin. The outer graben overlap the radial distance range of wrinkle ridges (Fig. 4). The basin-radial graben include more than 230 linear troughs that converge near the center of the basin (~30°N, 163°E) as part of Pantheon Fossae, some of whose distal portions were previously mapped as part of the polygonal graben patterns in the Mariner 10 images of eastern Caloris (*10*, *11*, *13*). The measured lengths of individual graben range from ~5 to ~110 km, and widths range from less than 1 km up to a maximum of ~8 km. An impact crater about 40 km in diameter, Apollodorus, is located near the center of the complex (black arrow, Fig. 1, A and B). The crater's rim, wall, and floor expose dark material, as do other large craters superposed on the interior plains. The crater rim is not cross-cut by the graben, and proximal ejecta obscure the graben. These relations suggest that the graben postdate the plains and that the deposition of crater ejecta postdates the graben. Farther from the center of the basin, the polygonally arranged graben form a broad annulus of extensional deformation (Fig. 4). The distal graben of Pantheon Fossae appear as segments of the polygons and part of the regional extension. Wrinkle ridges extend farther from the basin center

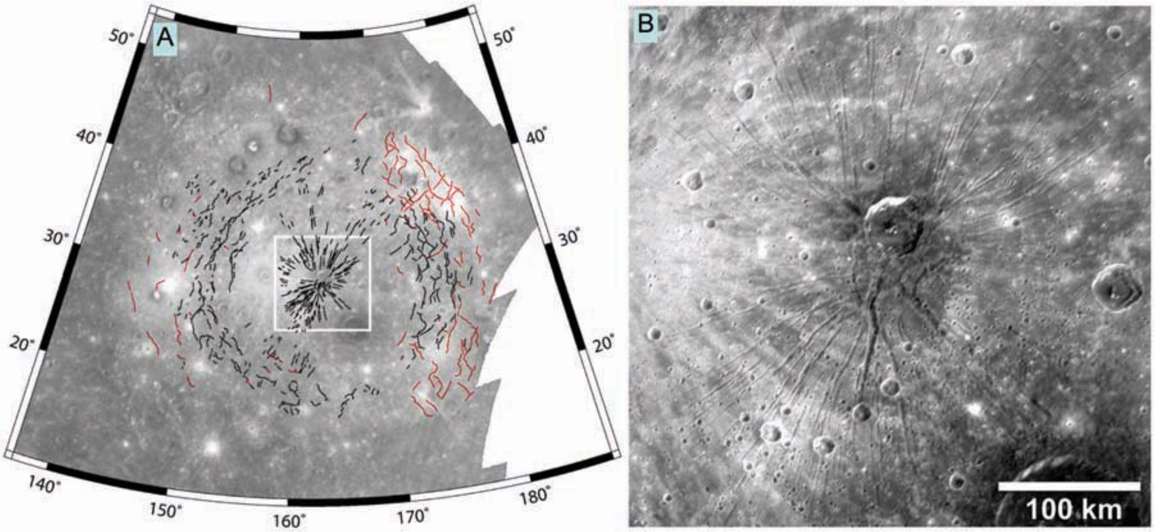


Fig. 4. (**A**) Lambert conformal tectonic map of the Caloris basin showing graben (black) and wrinkle ridges (red), overlaid on a high-resolution narrow-angle camera mosaic. The map is based on available images with the most favorable illumination for recognizing morphology (MESSENGER narrow-angle images in the central part of the basin and Mariner 10 images in the eastern part of the basin). (**B**) High-resolution image of the inset of the Pantheon Fossae, indicated by the white box in (A). North is up.

than do the graben. MESSENGER's high-Sun lighting geometry makes the recognition of low-relief ridges difficult; accordingly, a lower density of wrinkle ridges is recognized in MESSENGER imaging than in overlapping Mariner 10 imaging of eastern Caloris. The cross-cutting relations between wrinkle ridges and graben are consistent in both eastern and western Caloris; where the two types of features intersect, and their age relation can be determined, wrinkle ridges are always cut by and thus are older than the graben.

The pattern within Caloris of central radial graben and distal radial and concentric graben cross-cutting wrinkle ridges contrasts sharply with the spatial and temporal distribution of comparable features in lunar basins. In the lunar maria, wrinkle ridges occur predominantly in basin interiors, whereas graben are found in more distal parts or outside the basins (*22*, *23*). Wrinkle ridges occur in the youngest lunar mare basalts, whereas lunar graben are restricted to the oldest basalts (*24*, *25*). There is also no lunar counterpart to Pantheon Fossae. On the Moon, the distribution of wrinkle ridges and graben is thought to result from loading of the lithosphere by relatively dense, uncompensated mare basalt material, which induces subsidence and flexure of the lithosphere, leading to compression in the basin interiors and extension at the margins (*23*, *26*). The spatial and temporal distribution of tectonic features in Caloris cannot be fully explained by such models. Wrinkle ridges in Caloris may have formed in response to subsidence of the interior plains (*10*), possibly aided by a compressional stress due to global contraction (*11*).

Two models have been proposed to account for interior extension in Caloris: exterior loading and lateral crustal flow. In the exterior loading model, the superposition of smooth plains exterior to the Caloris basin results in an annular load that causes basin-interior uplift and extension (*10*, *12*). The lateral flow model involves movement of the lowest portions of postulated thick crust exterior to the basin inward toward the basin center, which causes late-stage uplift and extension (*11*). The central radial graben complex in Caloris presents a new constraint for these models, namely that tangential extensional stress exceeded radial extensional stress throughout the central basin region in order to account for the radial arrangement of the troughs. Another possibility is that Pantheon Fossae formed as the surface expressions of dikes that propagated radially from a magmatic intrusion near the basin center (*18*).

From the combination of MESSENGER and Mariner 10 data, six major phases in the history of the interior of the Caloris basin can be recognized: (i) formation of an impact basin and emplacement of impact melt (*27*) that must now be buried; (ii) superposition of several large craters in the basin interior; (iii) volcanic emplacement of light, redder material to form the interior plains, in part by pyroclastic and in part by effusive processes; (iv) formation (or continued activation) of wrinkle ridges in the outer part of the basin; (v) formation of radial and concentric graben; and (vi) superposition of still more impact craters, with smaller craters penetrating and excavating only the youngest volcanic plains and larger ones penetrating through to underlying darker material. Analysis of crater density variations (*28*) suggests that the emplacement of smooth plains exterior to the basin continued after phase (iii).

References and Notes

1. S. C. Solomon *et al.*, *Science* **321**, 59 (2008).
2. P. D. Spudis, J. E. Guest, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 118–164.
3. J. E. Guest, R. Greeley, U.S. Geological Survey (USGS) Miscellaneous Investigation Series, Map I-1408 (USGS, Denver, CO, 1983).
4. G. G. Schaber, J. F. McCauley, USGS, Map I-1199 (USGS, Denver, CO, 1980).
5. J. W. Head *et al.*, *Space Sci. Rev.* **131**, 41 (2007).
6. J. K. Harmon *et al.*, *Icarus* **187**, 374 (2007).
7. J. F. McCauley, *Phys. Earth Planet. Inter.* **15**, 220 (1977).
8. N. J. Trask, R. G. Strom, *Icarus* **28**, 559 (1976).
9. D. E. Wilhelms, *Icarus* **28**, 551 (1976).
10. H. J. Melosh, W. B. McKinnon, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ. 1988), pp. 374–400.
11. T. R. Watters, F. Nimmo, M. S. Robinson, *Geology* **33**, 669 (2005).
12. P. J. Kennedy, A. M. Freed, S. C. Solomon, *J. Geophys. Res.*, in press, 10.1029/2007JE002992 (2008).
13. R. G. Strom, N. J. Trask, J. E. Guest, *J. Geophys. Res.* **80**, 2478 (1975).
14. T. R. Watters, *J. Geophys. Res.* **93**, 10236 (1988).
15. T. R. Watters, *Icarus* **171**, 284 (2004).
16. S. E. Hawkins III *et al.*, *Space Sci. Rev.* **131**, 247 (2007).
17. M. S. Robinson *et al.*, *Science* **321**, 66 (2008).
18. J. W. Head *et al.*, *Science* **321**, 69 (2008).
19. C. M. Pieters, in *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, C. Pieters, P. Englert, Eds. (Cambridge Univ. Press, New York, 1993), pp. 309–339.
20. J. W. Head, A. Gifford, *Moon Planets* **22**, 235 (1980).
21. J. B. Adams, C. Pieters, T. B. McCord, *Proc. Lunar Sci. Conf.* **5**, 171 (1974).
22. M. P. Golombek, *J. Geophys. Res.* **84**, 4657 (1979).
23. S. C. Solomon, J. W. Head, *Rev. Geophys. Space Phys.* **18**, 107 (1980).
24. B. K. Lucchitta, J. A. Watkins, *Proc. Lunar Planet. Sci. Conf.* **9**, 3459 (1978).
25. J. M. Boyce, *Proc. Lunar Science Conf.* **7**, 2717 (1976).
26. H. J. Melosh, *Proc. Lunar Planet. Sci. Conf.* **9**, 3513 (1978).
27. M. J. Cintala, *J. Geophys. Res.* **97**, 947 (1992).
28. R. G. Strom *et al.*, *Science* **321**, 79 (2008).
29. N. Laslo, H. Kang, R. Vaughan, A. Harch, R. Shelton, and A. Berman designed the imaging sequences that made this contribution possible. B. Denevi, K. Becker, and C. Hash are gratefully acknowledged for data calibration and processing. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to the Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington.

17 April 2008; accepted 3 June 2008
10.1126/science.1159261

REPORT

Laser Altimeter Observations from MESSENGER's First Mercury Flyby

Maria T. Zuber,[1]* David E. Smith,[2] Sean C. Solomon,[3] Roger J. Phillips,[4] Stanton J. Peale,[5] James W. Head III,[6] Steven A. Hauck II,[7] Ralph L. McNutt Jr.,[8] Jürgen Oberst,[9] Gregory A. Neumann,[2] Frank G. Lemoine,[2] Xiaoli Sun,[2] Olivier Barnouin-Jha,[8] John K. Harmon[10]

A 3200-kilometers-long profile of Mercury by the Mercury Laser Altimeter on the MESSENGER spacecraft spans ~20% of the near-equatorial region of the planet. Topography along the profile is characterized by a 5.2-kilometer dynamic range and 930-meter root-mean-square roughness. At long wavelengths, topography slopes eastward by 0.02°, implying a variation of equatorial shape that is at least partially compensated. Sampled craters on Mercury are shallower than their counterparts on the Moon, at least in part the result of Mercury's higher gravity. Crater floors vary in roughness and slope, implying complex modification over a range of length scales.

Topography is a fundamental measurement to characterize quantitatively the surfaces of solid planetary bodies at length scales ranging from the long-wavelength shape to such local and regional processes as impact cratering, volcanism, and faulting. During the first flyby of Mercury by the MESSENGER spacecraft on 14 January 2008 (*1*), the Mercury Laser Altimeter (MLA) (*2*, *3*) successfully ranged to the planet's surface, providing the first altimetric observations of the planet from a spacecraft.

Previous measurements of the shape and topography of Mercury had been derived from Earth-based radar ranging (*4*, *5*) constrained by range observations from Mariner 10 (*6*). Because of the low inclination (7°) of Mercury's orbital plane to the ecliptic, Earth-based altimetric profiles are limited to ±12° latitude and have a spatial resolution of ~6×100 km^2 and a vertical precision of 100 m (*5*). These observations indicated a planetary reference radius of 2440 ± 1 km, an equatorial ellipticity of 540 ± 54 $\times 10^{-6}$, and an equatorial center of figure (COF) offset from the planet's center of mass (COM) of 640 ± 78 m in the direction 319.5° ± 6.9° W (*6*, *7*).

The MLA profile (Fig. 1) was acquired approximately along Mercury's equator, in a region that was in darkness during the flyby, and within the hemisphere not imaged by Mariner 10. Consequently, there are no optical images of the region in which altimetry was collected, so we used an Arecibo radar image (*8*) to correlate the profile with surface features. The MLA began ranging ~1 min before the spacecraft's closest approach and continued for ~10 min. Usable returns were received up to an altitude of 1500 km, which was larger than the expected maximum of 1200 km (*2*). As the spacecraft velocity and range from Mercury changed during the flyby, the size of laser spots on the surface varied from 23 to 134 m and the shot spacing varied from 888 to 725 m (*9*). The vertical precision varied with the received signal strength and is <15 cm at the closest range, limited by the resolution of the timing electronics. The radial accuracy of ~100 m is limited by uncertainties in the trajectory associated with errors in the ephemerides of MESSENGER and Mercury.

The profile spans ~ 20% of the circumference of the planet and shows a 5.2-km dynamic range of topography and 930-m root-mean-square (RMS) roughness (Fig. 1). The radius of Mercury apparently decreases by 1.4 km along the equator from ~10° to 90° E, corresponding to a 0.02° downward slope to the east. This long-wavelength surface tilt begins 30° west of the previously estimated COF/COM offset (*6*) and was not sampled in Earth-based radar altimetry (*4*). Such a long-wavelength slope, if a fundamental feature of the equatorial shape of the planet, might arise from crustal thickness or crustal density variations, global-scale mantle density variations, or topography along the planet's core-mantle boundary, which for Mercury is ~600 km beneath the surface.

The slope can be interpreted in the context of an ellipsoidal planetary shape (*10*). If we suppose that the difference in principal moments of inertia, $B - A$, is entirely a result of an ellipsoidal distribution of surface mass with density ρ_s and with semi-axes $a > b > c$, then

$$B - A = \frac{4\pi\rho_s abc}{15}\left(a^2 - b^2\right) \approx \frac{8\pi\rho_s R^4}{15}(a - b) \quad (1)$$

from which we may write

$$a - b = \frac{5}{2}R\left(\frac{B - A}{C_m}\right)\left(\frac{C_m}{C}\right)\left(\frac{C}{MR^2}\right)\frac{\langle\rho\rangle}{\rho_s} \quad (2)$$

where $A < B < C$ are the principal moments of inertia of Mercury, C_m is the moment of inertia of the mantle and crust alone, and M,

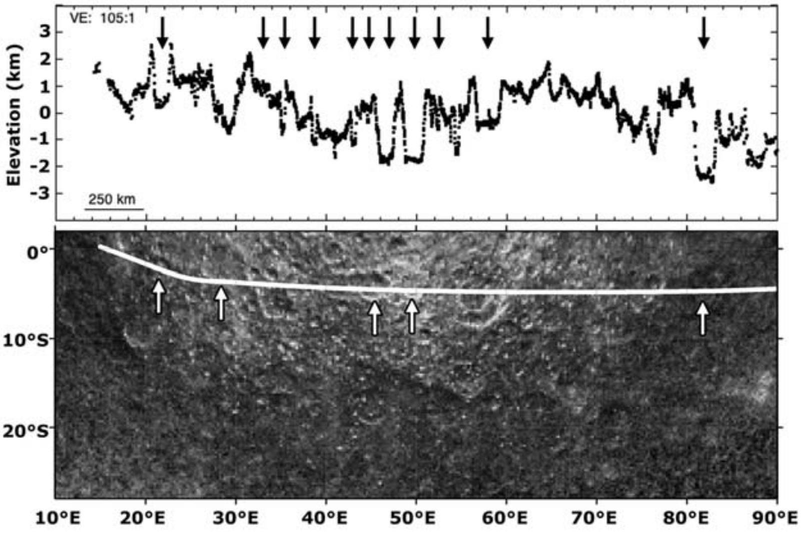


Fig. 1. (**Top**) MLA profile (vertical exaggeration 105:1). (**Bottom**) Arecibo radar image [adapted from (*8*)] with MLA profile location (white line) superposed. Arrows at top indicate locations of craters in Table 1 interpreted from detailed analysis of MLA profile points. The locations of several of the major craters are indicated by arrows on the radar image. The two-ringed circular structure in the Arecibo image at ~55 to 60°E is represented in part by a deep depression in the altimetry, but north-south radar ambiguities may be contributing to the structure in the image.

[1] Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139–4307, USA. [2] Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [3] Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA. [4] Southwest Research Institute, Boulder, CO 80302, USA. [5] Department of Physics, University of California, Santa Barbara, CA 93106, USA. [6] Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [7] Department of Geological Sciences, Case Western Reserve University, Cleveland, OH 44106, USA. [8] Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [9] Institute of Planetary Research, German Aerospace Center (DLR), Berlin, D-12489 Germany. [10] National Astronomy and Ionosphere Center, Arecibo Observatory, Arecibo 00612, Puerto Rico.

*To whom correspondence should be addressed. E-mail: zuber@mit.edu

Table 1. Crater apparent slopes, RMS roughness, and pulse widths.

Longitude (°E)	Apparent floor diameter* (km)	Range (km)	Emission angle† (deg)	Apparent floor slope‡ (deg)	RMS roughness (m)	Pulse width (ns)	Spot roughness§ (m)
21.1–22.4	55.3	415	61	0.298	89.8	5.0 ± 2.1	0.3 ± 0.7
33.7–33.9	8.5	293	49	–0.451	11.8	12.4 ± 1.8	1.8 ± 3.4
35–35.3	12.8	289	48	1.26	24.3	14.6 ± 10.8	2.3 ± 2.4
38.5–38.9	17	285	44	–0.571	46.3	22.5 ± 18.4	3.8 ± 3.9
42.9–43.2	12.8	293	40	0.252	20.3	12.0 ± 10.6	1.5 ± 2.1
44.5–44.6	4.3	298	38	–0.573	7.9	19.6 ± 20.7	2.9 ± 4.0
46–47.3‖	55.4	313	36	0.177	66.5	19.3 ± 14.5	2.8 ± 2.7
48.8–50.6‖	76.7	336	34	–0.0784	31.9	20.9 ± 15.3	3.0 ± 2.7
52.3–52.4	4.3	365	31	0.117	5.7	18.0 ± 16.4	2.4 ± 2.9
56.8–59.1	98	429	26	0.0852	96.5	20.6 ± 16.7	2.7 ± 2.8
81–82.3	55.4	1020	1.5	–0.0008	110	6.3 ± 2.1	0.2 ± 0.3

*Length of MLA-crossing chord. †Emission angle is the angle between the range vector and surface normal. ‡Positive slopes are defined to be downward to the west. §Roughness from pulse width corrected for spacecraft range and emission angle. ‖Craters shown in Fig. 2.

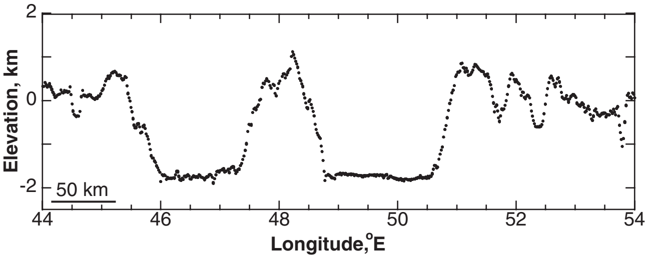


Fig. 2. Close-up of two craters showing contrast in floor roughness and tilt. The vertical exaggeration is 30:1.

R, and $\langle\rho\rangle$ are the mass, radius, and mean density of Mercury, respectively. The form of the right hand side of Eq. (2) is convenient because from measurements we have $(B - A)/C_m = (2.03 \pm 0.12) \times 10^{-4}$ (*11*), and from models we have C_m/C = 0.4 to 0.7 and C/MR^2 = 0.31 to 0.35 (*12*, *13*). The value of $(a - b)$ from MLA is 1.4 km. Surface shell densities of 2000 and 3000 kg m^{-3}, which bracket likely values, yield $(a - b)$ values of 0.26 to 0.87 km. These $(a - b)$ values are less than that observed, implying that the surface topography is at least partially compensated. The simplest explanation is that support of topography occurs by variations in crustal thickness, an inference that has also been invoked to explain Mercury's COF-COM offset by analogy with the situation on other terrestrial planets (*6*).

MLA profiled numerous depressions interpreted as impact craters on the basis of topographic expression and appearance on the Arecibo image. As on other terrestrial planets, the geomorphological complexity of impact craters on Mercury increases with diameter (*14*), with craters undergoing a transition at a diameter of about 11 km from a simple bowl shape to a planform with a flat floor, slumped walls, and a central peak (*14*). On a given planet, the ratio of depth to diameter (d/D) is uniform for unmodified complex craters, and where the MLA profile crossed close to crater centers, the ratio is ~1/40, less than on the Moon (d/D ~ 1/20). Two examples are craters in the longitude range ~45° to 50° E that have diameters (107 km and 122 km) comparable to that of Tycho (102 km), among the largest fresh craters on the Moon's nearside (Fig. 2). Whereas Tycho has a depth of 4.8 km (*15*), these craters have depths of 2.4 and 2.9 km, respectively. Although these craters may have undergone postformation modification, their substantially shallower depths in comparison to lunar counterparts is likely due at least in part to Mercury's higher surface gravity (*16*).

Crater floors may preserve evidence for modification processes that bear on geological evolution. From MLA we characterized the floors of complex craters by measuring apparent (along-track) slope, RMS roughness, and the widths of returned laser pulses, the last of which are indicative of topographic variance (due to roughness and footprint-scale slopes) within individual laser spots (*17*). Along-track slopes of 11 crater floors range from –10 m km^{-1} to +22 m km^{-1} (–0.57° to +1.26°) (*18*) (Table 1) and do not display an obvious pattern; most notably, these floor slopes do not correlate with the eastward long-wavelength slope. The RMS roughness over the approximate length scale of the crater diameter ranges from 5.7 to 110 m. Pulse widths vary considerably within individual craters, from 6 to >60 ns, indicating 2 to >20 m of vertical variability on horizontal scales of tens to hundreds of meters. For the craters studied, apparent slope, RMS roughness, and pulse widths are uncorrelated, which implies that the processes that caused tilting and created the roughness of crater floors are complex and do not operate uniformly over different length scales. Potential sources of modification include anelastic relaxation, volcanic resurfacing, tectonic subsidence, wall slumping, and ejecta emplacement from younger nearby impacts, and the variability implies that a combination of these processes operated on the profiled craters.

References and Notes

1. S. C. Solomon *et al.*, *Science* **321**, 59 (2008).
2. J. F. Cavanaugh *et al.*, *Space Sci. Rev.* **131**, 451 (2007).
3. The MLA is a time-of-flight laser rangefinder that uses direct detection and pulse-edge timing to determine precisely the range from the MESSENGER spacecraft to the target surface. MLA's laser transmitter emits 5-ns-wide pulses at an 8-Hz rate with 20 mJ of energy at a wavelength of 1064 nm. Return echoes are collected by an array of four refractive telescopes that are detected using a single silicon avalanche photodiode detector. The timing of laser pulses is measured using a set of time-to-digital converters and counters and a crystal oscillator whose frequency was monitored periodically from Earth.
4. J. K. Harmon, D. B. Campbell, D. L. Bindschadler, J. W. Head, I. I. Shapiro, *J. Geophys. Res.* **91**, 385 (1986).
5. J. K. Harmon, D. B. Campbell, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Ariz. Press, Tucson, AZ, 1988), pp. 101–117.
6. J. D. Anderson, R. F. Jurgens, E. L. Lau, M. A. Slade III, G. Schubert, *Icarus* **124**, 690 (1996).
7. C. F. Yoder, in *Global Earth Physics, A Handbook of Physical Constants*, T. J. Ahrens, Ed. (American Geophysical Union, Washington, DC, 1995).
8. J. K. Harmon *et al.*, *Icarus* **187**, 374 (2007).
9. The profile in Fig. 1 was computed using a reconstructed trajectory, msgr_20040803_20120401_od118.bsp, supplied by the MESSENGER navigation team.
10. In this configuration, 0° longitude coincides with one of the hot poles of Mercury, which is on the axis of minimum moment of inertia and 90° on the equator is in the direction of the intermediate axis of inertia. The radius of the ellipsoid equals *a* at 0° E and decreases to *b* at 90° E.
11. J.-L. Margot, S. J. Peale, R. F. Jurgens, M. A. Slade, I. V. Holin, *Science* **316**, 710 (2007).
12. H. Harder, G. Schubert, *Icarus* **151**, 118 (2001).
13. S. A. Hauck II, S. C. Solomon, D. A. Smith, *Geophys. Res. Lett.* **34**, L18201, 10.1029/2007GL030793 (2007).

14. R. J. Pike, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 165–274.
15. J.-L. Margot, D. B. Campbell, R. F. Jurgens, M. A. Slade, *J. Geophys. Res.* **104**, 11875 (1999).
16. R. J. Pike, *Proc. Lunar Planet. Sci. Conf.* **11**, 2159 (1980).
17. Surface roughness and small-scale tilts, and off-nadir ranging, which occurred at angles of up to 70° during the flyby, combine to spread MLA's reflected pulse. To maximize the probability of detecting spread returns, the returned pulse is passed through three matched filters after the detector amplifier, and subsequently time-to-digital converters are employed to measure the leading- and trailing-edge times of the pulse. MLA measures pulse widths between 6 and 1000 ns, corresponding to an RMS variation in range to the surface within each laser spot of 0.4 to 64 m (*2*, *19*).
18. These apparent slopes are likely minimum values.
19. J. B. Abshire, X. Sun, R. S. Afzal, *Appl. Opt.* **39**, 2449 (2000).

20. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington. We acknowledge important contributions from the MLA instrument and MESSENGER spacecraft teams.

14 April 2008; accepted 3 June 2008
10.1126/science.1159086

REPORT

Mercury Cratering Record Viewed from MESSENGER's First Flyby

Robert G. Strom,[1]* Clark R. Chapman,[2] William J. Merline,[2] Sean C. Solomon,[3] James W. Head III[4]

Morphologies and size-frequency distributions of impact craters on Mercury imaged during MESSENGER's first flyby elucidate the planet's geological history. Plains interior to the Caloris basin displaying color and albedo contrasts have comparable crater densities and therefore similar ages. Smooth plains exterior to Caloris exhibit a crater density ~40% less than on interior plains and are thus volcanic and not Caloris impact ejecta. The size distribution of smooth-plains craters matches that of lunar craters postdating the Late Heavy Bombardment, implying that the plains formed no earlier than 3.8 billion years ago (Ga). At diameters less than or equal to 8 to 10 kilometers, secondary impact craters on Mercury are more abundant than primaries; this transition diameter is much larger than that on the Moon or Mars. A low density of craters on the peak-ring basin Raditladi implies that it may be younger than 1 Ga.

Mercury has been struck by asteroids and comets since it formed, resulting not only in primary impact craters of all sizes but also in secondary craters made by reimpact of ejecta from the primary craters. Such secondaries typically have morphologies different from the pristine shapes of primary craters, and many secondary craters form clusters and chains. Geological processes such as faulting, volcanism, downslope motion, and continued cratering all degrade crater shapes, eventually erasing them by erosion or covering. The statistics of crater sizes, shapes, and spatial relations—especially their size-frequency distributions (SFDs) (*1*)—provide information (including relative ages) about the processes that formed and reshaped the cratered landscapes.

Images of Mercury by Mariner 10 from 1974 to 1975 and subsequent studies of other planetary surfaces have raised issues that the MESSENGER mission to Mercury can address (*2*), including the relative importance of secondary versus primary cratering and of volcanic versus impact-ejecta modes of plains formation. During its first flyby of Mercury, the MESSENGER spacecraft imaged portions of a crater-scarred landscape never before seen at close range. These images show the entire 1550-km-diameter Caloris impact basin (*3*) and a broad surrounding annulus of smooth plains. They also reveal diverse cratered terrains, some nearly saturated with large craters but others very sparsely cratered. Here we report preliminary analyses of crater morphology and SFD measurements from several selected regions, based chiefly on images from the narrow-angle camera (NAC) of the Mercury Dual Imaging System (*4*).

Observations of the Moon and Mars have shown that craters in the inner solar system have two SFD components (*5*, *6*). Terrains with a high density of large craters have a complex differential SFD that approximately follows a power law with a slope of –2 for crater diameter $D = {\sim}2$ to 50 km. This "Population 1" was formed primarily during the Late Heavy Bombardment (LHB), characterized by large impact basins such as the Caloris basin on Mercury, only part of which was seen by Mariner 10. Younger surfaces have an SFD with a slope of –3 ("Population 2"). (Figure S1 summarizes these two SFDs.) Both impactor populations were probably derived originally from the asteroid belt (*6*). Population 1 may have resulted from size-independent ejection as gravitational resonances swept through the belt during giant-planet migration ~4 billion years ago (Ga) (*7*, *8*). Population 2 reflects impacts of near-Earth asteroids (NEAs), mainly derived by the size-dependent Yarkovsky effect that causes smaller main-belt asteroids to preferentially enter resonances and be placed into planet-crossing orbits (*9*, *10*). Heavily cratered regions of Mercury have a SFD similar to that of the highlands of the Moon and Mars (Fig. 1A). However, on Mercury and Mars, there is a dearth of craters with $D < 40$ km relative to the Moon (*11*). On Mercury, smaller craters were apparently removed by the formation of "intercrater plains" (*12*) during the LHB (different processes erased smaller Martian craters). The shapes of the SFDs for Mercury, Mars, and the Moon for $D = 40$ to 150 km (a range not affected by intercrater plains and with good statistics) match each other better if shifted somewhat in diameter, apparently because heliocentrically orbiting NEAs collide with Mercury at higher velocities and Mars at lower velocities as compared with the Moon (fig. S4).

MESSENGER data show that the northwestern half of the Caloris floor has a slightly lower albedo and different color than the southeastern half (Fig. 1C) (*3*). Could these two plains regions have been formed by volcanic episodes at widely different times? SFDs (for $D > 10$ km, craters unlikely to be secondaries) for these regions (Fig. 1B) (as well as for an east/west division) show no significant differences, so their ages are comparable within 10 to 20%. Until we can measure crater densities on the Caloris rims and ejecta, we cannot determine if the flooding of the floor was contemporaneous with the impact (e.g., by impact melt) or occurred later; the uncertainties nonetheless permit a lengthy period for emplacement of successive volcanic flows, particularly if they occurred after the end of the LHB when the cratering rate was low.

Since the Apollo-era discovery that the Cayley plains on the Moon were basin impact ejecta rather than of volcanic origin, a major issue in planetary geology has concerned the relative importance of volcanism in plains formation. The crater density on some of the darker exterior smooth plains that form an annulus around Caloris is about 40% lower than that on plains inside Caloris (Fig. 2A), which is consistent with

[1]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [2]Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, CO 80302, USA. [3]Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington, DC 20015, USA. [4]Department of Geological Sciences, Brown University, Providence, RI 02912, USA.

*To whom correspondence should be addressed. E-mail: rstrom@lpl.arizona.edu

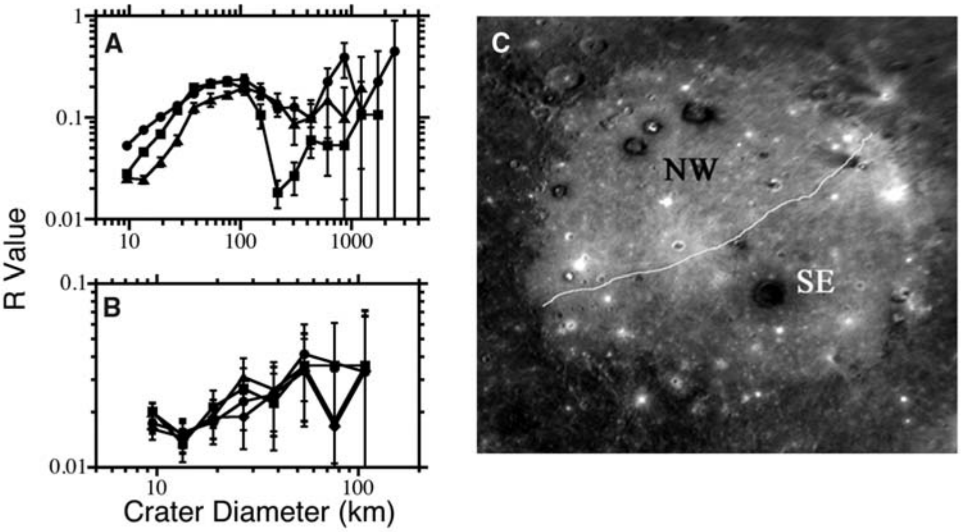


Fig. 1. (**A**) Crater SFDs for heavily cratered surfaces on the Moon (circles), Mars (squares), and Mercury (triangles) [from (*11*)]. Mars and Mercury have a deficit of craters at $D < 40$ km. The upturn at 10 km for Mercury could be the start of the secondary branch (Fig. 4). The format of this diagram (and of others in this paper) is the log-log R plot, a version of the differential SFD (*1*, *24*). Error bars in all figures indicate $\sqrt{(N)}/N$ SDs and do not account for uncertain systematic errors. (**B**) Crater SFDs of the northwestern (triangles) and southeastern (diamonds) portions of the Caloris interior plains, and separately of the eastern (circles) and western (squares) halves of the basin. Crater densities for all four sectors are the same to within 10 to 20%. (**C**) MESSENGER image of the Caloris basin showing the regions of the interior plains having slightly different colors and albedos (*3*) for which craters were counted.

a Mariner 10 study in a smaller area (*13*). Thus, the exterior smooth plains were emplaced after the interior plains and must be volcanic units rather than ejecta deposits from the Caloris impact. The SFDs for the exterior and interior plains have shallower slopes than the SFDs for heavily cratered terrains, so the craters must be largely Population 2. Although the smooth plains contain some slightly degraded craters, most are fresh with well-defined ejecta deposits and have a Population 2 SFD like fresh lunar craters (Fig. 2B). Hence, the exterior plains formed near or after the end of the LHB at ~3.8 Ga and well after Caloris (*14*).

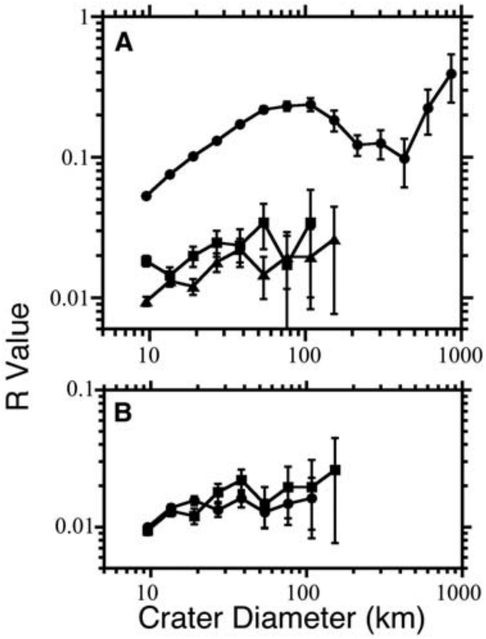


Fig. 2. Comparison of SFDs for the Caloris interior and exterior plains (Fig. 1B) as well as lunar highlands and fresh craters. (**A**) Circles represent the lunar highlands. The Caloris interior plains (squares) show a ~40% greater crater density than that of the exterior plains (triangles). (**B**) SFD for Caloris exterior plains (squares) slopes gently down to the left but is similar to the SFD for fresh lunar frontside craters (circles), indicating that the exterior (and interior) plains consist mainly of Population 2 craters.

The issue of whether secondaries are a minor or dominant contributor to small crater populations on planets and satellites has been controversial since the 1960s. On the basis of the steeply sloped SFD of lunar craters <2 km in diameter (resembling the D^{-5} SFD for secondaries of the lunar crater Langrenus), Shoemaker originally hypothesized that most small lunar craters are secondaries (*15*, *16*). More recently, it has been proposed (*17*) that this steep branch of the SFD is an inherent part of the primary crater SFD due to NEAs, but small secondaries have been shown to predominate on bodies as diverse as Europa and Mars (*18*–*20*). The degree to which secondary craters dominate crater populations remains controversial.

Clusters of secondaries, seen in some higher-resolution Mariner 10 images (*21*), were presumed to constitute a minor fraction of Mercury's smaller craters. MESSENGER images suggest that secondary cratering is much more important than had been thought, as exemplified by the many distinct chains and clusters of craters radiating away from prominent, large, fresh impact craters and basins. We characterized the secondary crater SFD for Mercury from craters within the overlapping secondary crater fields of three large, fresh primary craters and one smaller, rayed crater. The SFD slope (open diamonds in Fig. 3) for $4 < D < 10$ km is steep (at least –5) like that for lunar secondaries. More than half of these craters are in obvious clusters or chains, and >90% of craters with $2 < D < 4$ km are clustered; a major fraction of these must be secondaries because primary craters are spatially random. Also, ~90% of craters with $D < 10$ km have nonpristine shapes, characteristic of secondary craters. The actual proportion of secondaries in this region could be still higher, because high-velocity ejecta from more distant primaries make secondaries that are less obviously clustered and are more nearly bowl-shaped. SFDs for the other localities plotted in Fig. 3 (all generally far from obvious secondary crater fields) are also steep for $D < 8$ km (although steeper for the heavily cratered study area than for the plains west of Caloris), suggesting that secondaries dominate the small crater populations to varying degrees in these diverse terrains on Mercury.

Craters making up the steep secondary branch of SFDs begin to dominate over the primary SFD at a much larger diameter on Mercury than on the Moon, Mars, or other bodies (Fig. 4). The reason is not clear, but it was recognized from some Mariner 10 images (*22*) that secondaries on Mercury seem to be better preserved than those on the Moon. The density of impact basins on Mercury is about the same as on the Moon (Fig.

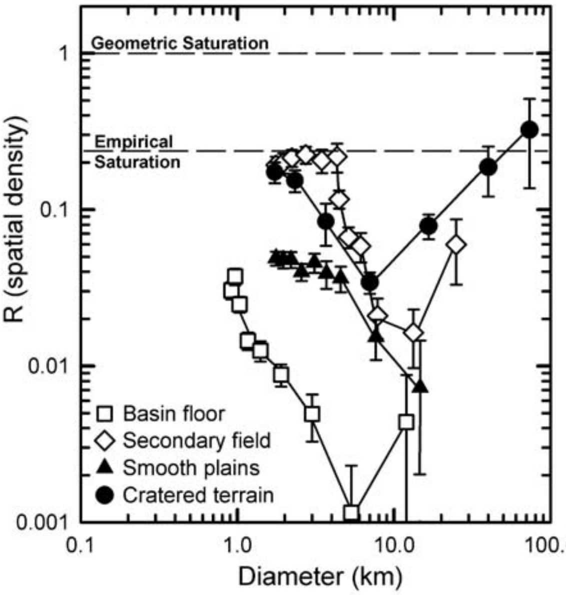


Fig. 3. SFDs for various surfaces on four frames from a NAC mosaic (frame locations are shown in fig. S3). The upturn at $D < \sim 8$ km is due to secondaries. For $2 < D < 4$ km in the secondary crater field, the SFD cannot continue upward to still higher density; consequently it must bend over to follow a –3 slope at the empirical saturation density. Note the very low density of craters on the floor of the peak-ring basin Raditladi (see text).

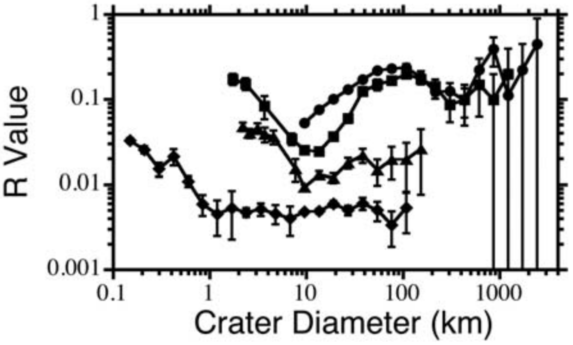


Fig. 4. Comparison of the lunar highlands (circles), heavily cratered terrain on Mercury (squares), Caloris exterior smooth plains (triangles), and young Martian surfaces (diamonds). The Mars SFD is a composite of young plains for $D > 0.2$ km. The upturn at $D < 1$ km is primarily attributable to secondaries (*20*). On Mercury, the upturn occurs at $D <$ 10 km and is due to secondaries. (On all terrestrial bodies, there may be a minority of craters >10 km that are secondaries from large basins.)

1A), so one might expect a comparable number of secondaries. Perhaps basins such as Caloris and the newly imaged peak-ring basin Raditladi are unusually youthful so that their secondaries are better preserved. Surface gravity cannot be responsible because it is similar on Mercury and Mars. Fragment size should be inversely proportional to impact velocity (*23*); thus, secondaries should be smaller on Mercury, where impact velocities are higher than those on the Moon and Mars. The larger secondaries may be the result of differences in material strength of the target material so that larger fragments are produced. In any event, the use of small craters for dating of geological units on Mercury must be done with even greater caution than is needed for other bodies. Whereas an older unit will tend to have more secondaries on it than a younger unit, there cannot be the one-to-one correspondence of crater density with relative or absolute age (as there is for primary craters) because of the temporally and spatially nonuniform production of secondaries.

The density of small craters on the floor of Raditladi is an order of magnitude lower than that of the plains west of Caloris, whereas the density on small craters on Raditladi's ejecta deposits is similar to that of its floor (Fig. 3). These results do not require that Raditladi is $^{1}/_{10}$ of the age of the plains, because it may have been formed when the cratering rate was changing rapidly as the LHB ended, compressing the time scale. Spatial densities of secondary craters vary, even across surfaces of the same age, but Raditladi is located near the apparently much older smooth plains that we sampled, so it is possible that Raditladi and its interior plains were formed within the last billion years. If the plains within Raditladi were formed by volcanic processes rather than being impact melt, then Mercury's geological activity may have persisted (at least locally) into comparatively recent epochs rather than ending shortly after the LHB.

References and Notes

1. In this paper, we display the crater SFDs using the "Relative" plot. For an R plot, the differential SFD is divided by $dN(D) \sim D^{-3}dD$ (where N is the number of craters within a diameter increment), because most crater SFDs are well described by a power law having an exponent within ±1 of –3. A –3 distribution plots as a horizontal line, a –2 distribution slopes down to the left at an angle of 45°, and a –4 distribution slopes down to the right at 45°. The vertical position measures crater density: The higher the position, the higher the crater density and the older the surface. Although ideally craters could be packed together so that spatial density $R = 1$, in reality cratered surfaces rarely exceed the empirical saturation density with $R \sim 0.2$ to 0.25.
2. J. W. Head *et al.*, *Space Sci. Rev.* **131**, 41 (2007).
3. S. L. Murchie *et al.*, *Science* **321**, 73 (2008).
4. S. E. Hawkins III *et al.*, *Space Sci. Rev.* **131**, 247 (2007).
5. R. G. Strom, S. K. Croft, N. G. Barlow, in *Mars*, H. H. Kieffer, B. M. Jakosky, C. Snyder, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1992), pp. 383–423.
6. R. G. Strom, R. Malhotra, T. Ito, F. Yoshida, D. A. Kring, *Science* **309**, 1847 (2005).
7. K. Tsiganis, R. Gomes, A. Morbidelli, H. G. Levison, *Nature* **435**, 459 (2005).
8. R. Gomes, H. F. Levison, K. Tsiganis, A. Morbidelli, *Nature* **435**, 466 (2005).
9. A. Morbidelli, D. Vokrouhlicky, *Icarus* **163**, 120 (2003).
10. W. F. Bottke, D. Vokrouhlicky, D. P. Rubincam, D. Nesvorny, *Annu. Rev. Earth Planet. Sci.* **34**, 157 (2006).
11. R. G. Strom, G. Neukum, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 336–373.
12. "Intercrater plains" are a type of terrain recognized on Mercury, the Moon, and other planets characterized by nearly level plains sprinkled with small craters interspersed among a moderate density of large craters. Their mode of formation has been controversial.
13. P. D. Spudis, J. E. Guest, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 118–164.
14. From Mariner 10 images, Neukum (*11*) estimated the age of the Caloris exterior plains at 3.85 Ga (near the end of the LHB), on the basis of a calibration with the lunar cratering record.
15. E. M. Shoemaker, in *Physics and Astronomy of the Moon*, Z. Kopal, Ed. (Academic Press, New York, 1962), pp. 283–359.
16. E. M. Shoemaker, in *The Nature of the Lunar Surface*, W. N. Hess, D. H. Menzel, J. A. O'Keefe, Eds. (Johns Hopkins Univ. Press, Baltimore, 1965), pp. 23–77.
17. G. Neukum, B. A. Ivanov, W. K. Hartmann, in *Chronology and Evolution of Mars* (Kluwer Academic, Dordrecht, Netherlands, 2001), pp 55–86.
18. E. B. Bierhaus, C. R. Chapman, W. J. Merline, *Nature* **437**, 1125 (2005).
19. A. S. McEwen, B. S. Preblich, E. P. Turtle, *Icarus* **176**, 351 (2005).
20. A. S. McEwen, E. B. Bierhaus, *Annu. Rev. Earth Planet. Sci.* **34**, 535 (2006).
21. D. E. Gault, J. E. Guest, J. B. Murray, D. Dzurisin, M. C. Malin, *J. Geophys. Res.* **80**, 2444 (1975).
22. D. H. Scott, *Phys. Earth Planet. Inter.* **15**, 173 (1977).
23. H. J. Melosh, *Geology* **13**, 144 (1985).
24. Crater Analysis Techniques Working Group, *Icarus* **37**, 467 (1979).
25. We thank M. Banks, who did many of the crater counts and helped with the analyses. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 (to the Johns Hopkins University Applied Physics Laboratory) and NASW-00002 (to the Carnegie Institution of Washington).

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/79/DC1
SOM Text
Figs. S1 to S4

18 April 2008; accepted 4 June 2008
10.1126/science.1159317

REPORT

The Structure of Mercury's Magnetic Field from MESSENGER's First Flyby

Brian J. Anderson,[1]* Mario H. Acuña,[2] Haje Korth,[1] Michael E. Purucker,[2]
Catherine L. Johnson,[3] James A. Slavin,[4] Sean C. Solomon,[5] Ralph L. McNutt Jr.[1]

During its first flyby of Mercury, the MESSENGER spacecraft measured the planet's near-equatorial magnetic field. The field strength is consistent to within an estimated uncertainty of 10% with that observed near the equator by Mariner 10. Centered dipole solutions yield a southward planetary moment of 230 to 290 nanotesla $R_M{}^3$ (where R_M is Mercury's mean radius) tilted between 5° and 12° from the rotation axis. Multipole solutions yield non-dipolar contributions of 22% to 52% of the dipole field magnitude. Magnetopause and tail currents account for part of the high-order field, and plasma pressure effects may explain the remainder, so that a pure centered dipole cannot be ruled out.

Of the terrestrial planets, only Earth and Mercury possess global magnetic fields. The Mariner 10 encounters with Mercury in 1974 and 1975 (*1*–*3*) yielded the surprising result that Mercury has a coherent, intrinsic magnetic field (*4*–*6*). Estimates for the planetary dipole moment derived from Mariner 10 data range from 170 to 350 nT $R_M{}^3$ (where R_M is Mercury's mean radius); the uncertainty arises from the difficulty of distinguishing dipole and quadrupole contributions (*7*). Here, we apply magnetic field observations from the first MESSENGER flyby of Mercury (*8*, *9*) on 14 January 2008, herein M1, to determine the planetary moment and test for higher-order terms and secular changes in the internal field.

The spacecraft was within Mercury's magnetosphere for ~30 min during the flyby (Fig. 1). The inbound bow shock crossing corresponded to an abrupt increase in field magnitude, and the inbound magnetopause was evident in the decreased 1- to 10-Hz fluctuations and reduced directional and magnitude variability. At point A, the field rotated anti-sunward, azimuth near 180°, and the polar angle increased indicating passage out of the tail plasma sheet into the southern lobe. At point B, the 1- to 10-Hz fluctuations increased, indicating a change in local plasma conditions, and the field began to rotate northward at the same time that the magnitude increased, suggesting entry into a region dominated by the intrinsic field, interpreted here as departure from the southern tail lobe. Two field depressions occurred between B and closest approach (CA) with little corresponding change in field direction. There was a drop in field magnitude at point C, again without a change in direction. The outbound magnetopause boundary was indicated by the onset of large variations in magnitude and direction. Step C is interpreted elsewhere as a signature of a double magnetopause (*10*). At the outbound bow shock, the magnitude returned to the preencounter level. The interplanetary magnetic field (IMF) before and after the flyby was predominantly anti-sunward. There is no evidence of substormlike behavior or energetic particles (*10*). Thus, the M1 observations provide measurements of the magnetic field through an entire transit near the equator for undisturbed conditions.

To assess Mercury's intrinsic field, we applied spherical harmonic analysis (SHA) to the M1 observations combined with those from the first and third Mariner 10 flybys, M10-I and M10-III, respectively. The contribution of the external field must be considered because the magnetospheric fields are comparable to the planetary field (*11*, *12*). We used three approaches to the external field: (i) to gauge the

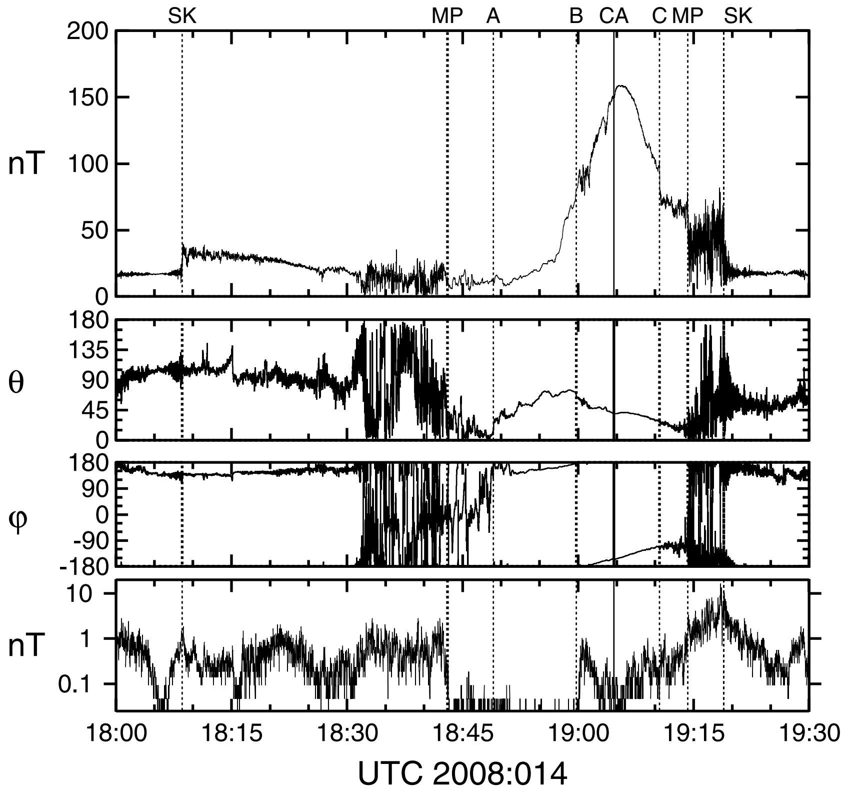


Fig. 1. MESSENGER Magnetometer observations for 14 January 2008 in polar Mercury solar orbital (MSO) coordinates versus UTC. θ is the polar angle (0° is north, normal to Mercury's orbital plane), and φ is azimuth (0° is sunward, 90° is duskward). Graphs show (**top**) field magnitude, (**middle**) θ and φ of the field direction, and (**bottom**) the 1- to 10-Hz bandpass fluctuation amplitude. Vertical lines indicate the bow shock crossings (SK), magnetopause crossings (MP), CA, and three transitions within Mercury's magnetosphere (A, B, and C). Data from B to C are used for intrinsic field analysis. Times (hh:mm:ss UTC) of the transitions are, for inbound SK, 18:08:38; inbound MP, 18:43:02; A, 18:52:04; B, 18:59:46; CA, 19:04:39; C, 19:10:34; outbound MP, 19:14:15; and outbound SK, 19:18:55.

[1] Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA. [2] Solar System Exploration Division, NASA Goddard Space Fight Center, Greenbelt, MD 20771, USA. [3] Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [4] Heliophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [5] Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA.

*To whom correspondence should be addressed. E-mail: brian.anderson@jhuapl.edu

significance of subsequent corrections, we first ignored the external field; (ii) we used the empirical Tsyganenko and Sitnov (TS04) model (*13*), developed for Earth's magnetosphere and adapted for Mercury to include magnetopause and tail currents (*11*, *12*); and (iii) we allowed external coefficients in the SHA to estimate the external field from the data under the assumption that the region sampled was current-free (*14*).

The IMF and solar wind pressure inputs for the TS04 model were determined separately for each flyby. The IMF was taken to be the average over ~10 min in the solar wind near the bow shock, judged to best represent conditions near CA. For M1 and M10-III, we used data after the outbound bow-shock crossing. For M10-I, it was likely that the IMF changed during the encounter (*11*), so we used data before the first inbound shock crossing. The solar wind pressure was determined by fitting the outbound magnetopause crossing for each encounter (*12*). For the M10-I and M10-III data, we used the same data intervals that have been analyzed previously (*3*).

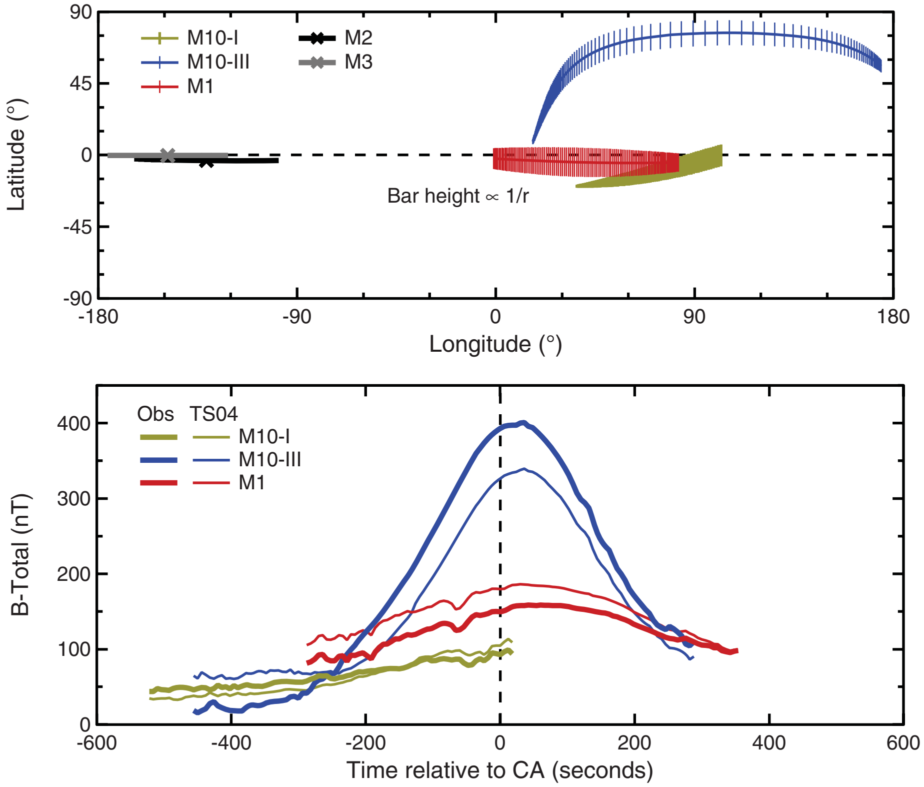


Fig. 2. Locations of all three encounters (**top**) and the field magnitude versus time relative to CA (**bottom**) for the intervals used in the analysis. The top graph shows latitude versus longitude for the M10-I, M10-III, and M1 encounters. Vertical bar heights are proportional to 1/*r*. The predicted latitude and longitude of the second and third MESSENGER flybys, M2 and M3, within 250 s of CA are also shown, where CA is indicated by crosses. In the bottom graph, thick lines show the raw data observations, and thin lines show observations with the TS04 external field subtracted.

Table 1. Dipole solutions to Mariner 10 and MESSENGER vector magnetic field data. A condition number (Cond. no.) lower than the ratio of the dipole moment to the root mean square deviation (RMSD) indicates a well-constrained solution. Δ_{xy} denotes the difference between solutions x and y; dipole magnitude (magn.) and direction (dir.).

Fit	External field approach	Data set(s)	Dipole (nT R_M^3)	Tilt	Longitude (north pole)	Condition number	RMSD (nT)
a	None	M10-I	−153	33°	−109°	2	15
b	None	M1	−168	17°	−135°	2	18
		Δ_{ab}	9% (magn.)		19° (dir.)		
c	TS04	M10-I	−197	18°	−29°	2	19
d	TS04	M1	−214	10°	−143°	2	11
		Δ_{cd}	8% (magn.)		24° (dir.)		
e	TS04	All	−229	9°	−161°	1	28
f	SHA 1 fit	All	−247	12°	−138°	10	26
g	SHA 3 fits	All	−290	5°	−161°	17	14
		Δ_{eg}	24% (magn.)		5.4° (dir.)		
		Δ_{fg}	16% (magn.)		7.1° (dir.)		

Determinations of the planetary field are influenced by the sampling coverage (Fig. 2). The M10-III observations were made near the northern pole, whereas M10-I and M1 measured the field near the equator. The maximum field magnitude recorded from M10-III was 401 nT, 2.5 times that observed from M1, 159 nT, even though CA for M10-III was at greater altitude than CA for M1, 327 km versus 201 km. This is not consistent with a centered dipole in vacuum because the maximum variation in field magnitude from a dipole is a factor of 2 at constant altitude.

The TS04-corrected values are more consistent with a dipole. The corrected field magnitude is lower for M10-III, higher for M1, and also higher for M10-I but only near CA, perhaps because the M10-I trajectory was at higher CA altitude (705 km). The magnetopause and tail currents reinforce the field over the pole and oppose the field near the equator on the nightside where CA for M1 and M10-I occurred, so correcting for the external field reduces the apparent pole-equator difference. The maximum magnitudes of the TS04-corrected fields are 339 nT for M10-III and 186 nT for M1. Scaling the value for M10-III from 330-km to 201-km altitude with a $1/r^3$ relation, where *r* is distance from the planet center, we estimated a maximum field at 201 km altitude near the north pole of 391 nT, 2.1 times greater than that observed during M1.

To assess whether the planetary moment changed since the Mariner 10 encounters, we compared observations from M10-I and M1, which followed similar trajectories. We evaluated the centered dipole separately from M10-I and M1 data with and without the TS04–external field correction (Table 1, fits a to d). The M10-I and M1 dipole results for the same external field correction agree to within 9% in magnitude and 24° in direction. Because results with and without the TS04–external field correction have comparable or greater differences, there is no statistically significant evidence for a change in the moment's magnitude or direction between 1974 and 2008.

Taking the intrinsic field to be unchanged, we used data from all three encounters to estimate the intrinsic field (fits e to g in Table 1). We calculated the SHA external field in two ways. First, we used data for the three encounters simultaneously to derive both the planetary centered dipole and a single set of external-field coefficients to degree and order 2 (SHA 1). We also derived separate external SHA fits to degree and order 2 for each encounter while solving for a single dipole internal field (SHA 3). These solutions yielded a planetary moment of 229 to 290 nT R_M^3, directed southward, with a tilt between 5° and 12° from the rotation axis.

Structure in the field from orders higher than a centered dipole was assessed in two ways. First, we estimated spherical harmonic coefficients through degree and order 2 (*7*, *14*). Second, we used a regularized, constrained solution (Reg. TS04) through degree and order 6 in the internal field with the TS04 external field to minimize the unknowns. This method solves for higher-order spherical harmonic coefficients while minimizing the power in non-dipole terms. It applies a smoothness constraint on the power spectrum of the harmonics (*15*, *16*) by using a formalism that allows a data misfit tolerance and a damping of the model structure (*17*). The degree-6 solution used sufficient terms to ensure that the results were controlled by the smoothness constraint rather than truncation.

The results (Table 2 and Fig. 3) suggest that the pole-equator magnitude difference may be the principal factor driving the quadrupole term in the solutions. The dipole models (Fig. 3) illustrate that a centered dipole cannot account for the observed range in field magnitudes. The quadrupole solutions, I, III, and IV, have essentially the same north-south asymmetry, dipole magnitudes ($B_{n=1}$ of 210 to 227 nT), and higher-order terms ($B_{n>1}$ of 93 to 103 nT). The solution with no correction for the external field is similar to the two SHA external-field solutions. The SHA method for the external field thus had relatively little effect on the intrinsic field solution. Although solution IV yields the lowest deviation, the condition number implies that it is less well constrained than solutions I, II, or III.

The north-south asymmetry is reduced in the two solutions derived from TS04-corrected data. Although the dipole fit to the TS04-corrected data still cannot account for the remaining pole-equator difference in the field, the higher-order TS04 fits yield smaller quadrupole terms. Solutions II and V have lower $B_{n=1}$ and $B_{n>1}$, and, although their deviations are higher than the other solutions, the condition number for II is low. The relative contribution of higher-order terms in the regularized solution is less than half that of solution IV.

The small equatorial field relative to that observed near the pole results in quadrupole magnitudes that are ~45% of the dipole. Applying an external field correction that includes MP and tail currents accounts for some of this difference, but the inferred intensity at 200 km from M10-III is still 50 nT too large relative to the field at the equator to be explained with a centered dipole. The corresponding equatorial field deficit is ~25 nT.

The TS04 external field correction does not account for the effects of local plasma pressures in Mercury's magnetosphere (*18*). Two depressions in field intensity between B and CA (Fig. 1) have signatures consistent with local plasma pressure, one from 19:00 to 19:02 universal time coordinated (UTC) and the second near 19:04 UTC. Both are associated with increases in the proton plasma count rates (*18*). Interpreted as plasma pressure signatures, these correspond to increases

Table 2. Magnetic field amplitudes due to dipole and higher-order terms from intrinsic field solutions to combined Mariner 10 and MESSENGER vector field data. Field amplitudes $B_{n=1}$ and $B_{n>1}$ are the square roots of the spherical harmonic power evaluated at 200-km altitude for $n = 1$ and $n > 1$, where n denotes the degree in the spherical harmonic solution. For solutions I through IV, $B_{n>1}$ is the $n = 2$ (quadrupole) amplitude. For solution V, $B_{n>1}$ is the square root of the summed spectral power for $n = 2$ to 6. Fit V used a data misfit tolerance of 1.2 and a structure damping parameter of 0.01 (*17*). The condition number is not applicable (n/a) for solution V. Spherical harmonic coefficients through degree 2 are listed in table S1, which also lists the norm for the terms of degree 3 through 6 of solution V.

	Fit: External field and harmonic analysis technique				
	I	II	III	IV	V
	None	TS04	SHA 1 fit	SHA 3 fits	Reg. TS04
$B_{n=1}$ (nT)	254	199	249	235	212
$B_{n>1}$ (nT)	124	87	130	118	46
$B_{n>1}/B_{n=1}$	0.49	0.44	0.52	0.50	0.22
RMSD (nT)	17	25	17	7	24
Condition number	8	8	12	50	n/a

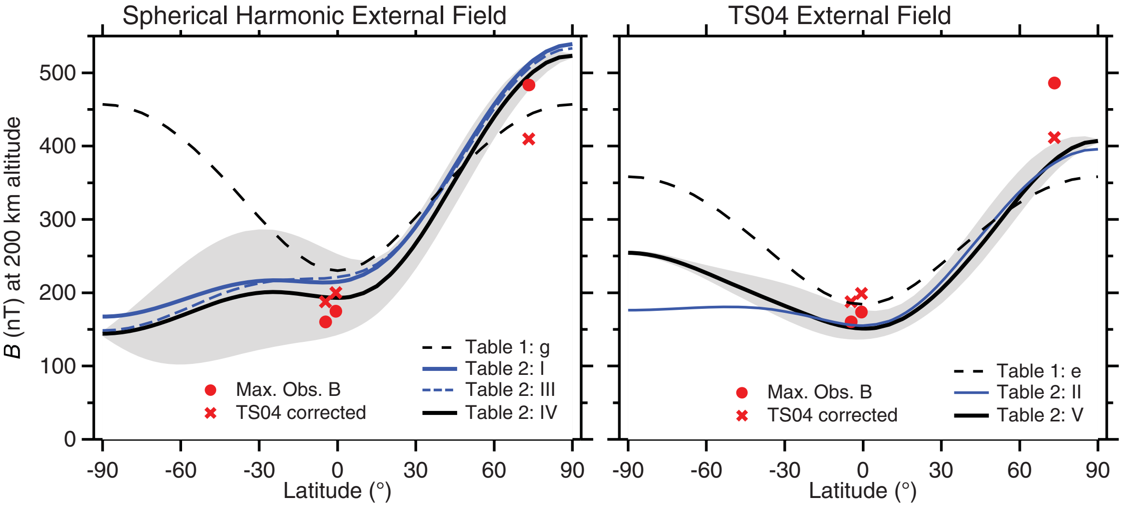


Fig. 3. Longitudinal averages versus latitude of 200-km-altitude magnetic field magnitudes from two classes of intrinsic field solutions as listed in Tables 1 and 2: those using either no external field or spherical harmonic representations for the external field (**left**) and those using the TS04 model adapted to Mercury (**right**). The observed magnitudes were mapped to 200 km with solution IV on the left and solution V on the right and by multiplying the observed magnitude by the ratio of the model field magnitude at 200 km and the observed altitude evaluated for the observation point latitude and longitude. As for the dipole inversions, we applied the SHA 1 and SHA 3 approaches to the spherical harmonic external field. In the left graph, the solid black line shows fit IV with gray shading indicating the range at a given latitude. Solid and dashed blue lines depict the results from fits I and III, respectively; and the dashed line shows the dipole fit g result. In the right graph, the solid black line shows fit V with gray shading as on the left, whereas the solid blue and dashed black lines show results from fits II and e, respectively. Red solid circles and crosses indicate the maximum magnetic field observed in each flyby extrapolated to 200 km.

in the plasma pressure of ~1.8 nPa. Also, coincident with the drop in field magnitude at C, the plasma proton count rates increased by a factor of three (*18*). The change in magnetic field magnitude implies a plasma pressure increase at C of ~2 nPa. Because the proton count rates before C were ~30% of those after C, the pressure before C was ~1 nPa, which would depress the field by ~7 nT.

Such signatures are consistent with hybrid simulations of Mercury's magnetosphere (*19*) that indicate an annulus of solar wind plasma within ~0.5 $R_{\rm M}$ altitude. The inward pressure gradient at the outer edge of such an annulus would suppress the magnetic field near the equator on the nightside and enhance it over the poles. The corresponding westward azimuthal current is about $I = hP/B$, where h is the vertical extent of the annulus, B is the magnetic field magnitude, P is the pressure in the annulus, and the pressure outside is taken to be zero. A 1-nPa pressure that goes to zero near 0.5 $R_{\rm M}$ altitude, where the field is ~50 nT, and that has a vertical extent of ~0.5 $R_{\rm M}$ corresponds to a current of 0.05 to 0.1 MA. This would decrease the equatorial field close to the planet by 10 to 30 nT and increase the field at the pole by ~5 to 10 nT. Thus, it is possible that the remaining deficit of equatorial field intensity of ~25 nT could be due to magnetospheric plasma. We conclude that an intrinsic quadrupole term is not required to account for the observations.

Recent simulations of Mercury's core dynamo suggest that the presence of a stagnant layer at the top of the molten outer core may suppress higher-order structure and yield secular variation over time scales of centuries rather than decades (*20*–*22*). We find no evidence for a change in the planetary dipole since 1974 and also find that the planetary field is predominantly and possibly entirely dipolar. Although there are significant uncertainties associated with these results, they are consistent with the presence of a stagnant outermost core.

References and Notes

1. N. F. Ness, K. W. Behannon, R. P. Lepping, Y. C. Whang, K. H. Schatten, *Science* **185**, 151 (1974).
2. N. F. Ness *et al.*, *J. Geophys. Res.* **80**, 2708 (1975).
3. N. F. Ness *et al.*, *Icarus* **28**, 479 (1976).
4. S. C. Solomon, *Icarus* **28**, 509 (1976).
5. L. J. Srnka, *Phys. Earth Planet. Inter.* **11**, 184 (1976).
6. D. J. Jackson, D. B. Beard, *J. Geophys. Res.* **82**, 2828 (1977).
7. J. E. P. Connerney, N. F. Ness, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, 1988), pp. 494–513.
8. S. C. Solomon *et al.*, *Science* **321**, 59 (2008).
9. B. J. Anderson *et al.*, *Space Sci. Rev.* **131**, 417 (2007).
10. J. A. Slavin *et al.*, *Science* **321**, 85 (2008).
11. J. G. Luhmann *et al.*, *J. Geophys. Res.* **103**, 9113 (1998).
12. H. Korth *et al.*, *Planet. Space Sci.* **52**, 733 (2004).
13. N. A. Tsyganenko, M. I. Sitnov, *J. Geophys. Res.* **110**, A03208, 10.1029/2004JA010798 (2005).
14. D. Winch, in *Encyclopedia of Geomagnetism and Paleomagnetism*, D. Gubbins, E. Herrero-Bervera, Eds. (Springer, Dordrecht, Netherlands, 2007), pp. 448–452.
15. R. L. Parker, *Geophysical Inverse Theory* (Princeton Univ. Press, Princeton, NJ, 1994).
16. J. D. Bloxham *et al.*, *Philos. Trans. R. Soc. London Ser. A* **329**, 415 (1989).
17. C. J. Johnson, C. G. Constable, *Geophys. J. Int.* **122**, 489 (1995).
18. T. H. Zurbuchen *et al.*, *Science* **321**, 90 (2008).
19. P. Trávníček *et al.*, *Geophys. Res. Lett.* **34**, L05104, 10.1029/2006GL028518 (2007).
20. U. R. Christensen, *Nature* **444**, 1056 (2006).
21. F. Takahashi, M. Matsushima, *Geophys. Res. Lett.* **33**, L10202 10.1029/2006GL025792 (2006).
22. J. Wicht *et al.*, *Space Sci. Rev.* **132**, 261 (2007).
23. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington. Support was also provided under the NASA MESSENGER Participating Science Program via grant NNX07AR73G.

14 April 2008; accepted 6 June 2008
10.1126/science.1159081

REPORT

Mercury's Magnetosphere After MESSENGER's First Flyby

James A. Slavin,[1]* Mario H. Acuña,[2] Brian J. Anderson,[3] Daniel N. Baker,[4] Mehdi Benna,[2] George Gloeckler,[5,6] Robert E. Gold,[3] George C. Ho,[3] Rosemary M. Killen,[6] Haje Korth,[3] Stamatios M. Krimigis,[3,7] Ralph L. McNutt Jr.,[3] Larry R. Nittler,[8] Jim M. Raines,[5] David Schriver,[9] Sean C. Solomon,[8] Richard D. Starr,[10] Pavel Trávníček,[11] Thomas H. Zurbuchen[5]

Observations by MESSENGER show that Mercury's magnetosphere is immersed in a comet-like cloud of planetary ions. The most abundant, Na$^+$, is broadly distributed but exhibits flux maxima in the magnetosheath, where the local plasma flow speed is high, and near the spacecraft's closest approach, where atmospheric density should peak. The magnetic field showed reconnection signatures in the form of flux transfer events, azimuthal rotations consistent with Kelvin-Helmholtz waves along the magnetopause, and extensive ultralow-frequency wave activity. Two outbound current sheet boundaries were observed, across which the magnetic field decreased in a manner suggestive of a double magnetopause. The separation of these current layers, comparable to the gyro-radius of a Na$^+$ pickup ion entering the magnetosphere after being accelerated in the magnetosheath, may indicate a planetary ion boundary layer.

The interaction of Mercury's magnetic field with the solar wind creates a small magnetosphere with a typical standoff altitude of ~0.5 $R_{\rm M}$ (where $R_{\rm M}$ is the mean planet radius; 1 $R_{\rm M}$ ~ 2440 km) (*1*, *2*) (Fig. 1). The MESSENGER spacecraft made the first of three flybys of Mercury on 14 January 2008 (*3*) and took measurements within Mercury's magnetosphere with its magnetometer (MAG) (*4*, *5*); energetic particle and plasma spectrometer, composed of the energetic particle spectrometer (EPS) and fast imaging plasma spectrometer (FIPS) (*6*, *7*); and x-ray spectrometer (XRS) (*8*).

The presence of the magnetosphere as an obstacle to the solar wind is signaled by the bow shock (BS), which was crossed at 18:08:38 (inbound) and 19:18:55 (outbound). Before the inbound magnetopause (MP) crossing at 18:43:02, the last extended interval of southward interplanetary magnetic field (IMF) ended at 18:38:40. The magnetosheath magnetic field was observed to be generally northward after the exit from the magnetosphere at 19:14:15. A northward IMF is unfavorable to dayside magnetic reconnection with Mercury's magnetic field and greatly limits the rate of solar wind energy transfer across the

[1]Heliophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [2]Solar System Exploration Division, NASA Goddard Space Fight Center, Greenbelt, MD 20771, USA. [3]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [4]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA. [5]Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, MI 48109, USA. [6]Department of Astronomy, University of Maryland, College Park, MD 20742, USA. [7]Academy of Athens, Athens 11527, Greece. [8]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA. [9]Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024, USA. [10]Department of Physics, Catholic University of America, Washington, DC 20064, USA. [11]Astronomical Institute, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

*To whom correspondence should be addressed. E-mail: james.a.slavin@nasa.gov

MP (2). The earlier southward IMF intervals before MESSENGER's entry into the magnetosphere were expected to produce strong energetic particle acceleration, as had been observed during Mariner 10's first flyby (2). The lack of measurable energetic electrons within the magnetosphere during MESSENGER's flyby (Fig. 2) indicates that energetic electrons remained within the magnetosphere for less than the ~4 min between the time when the southward IMF ended and when MESSENGER entered the magnetosphere.

MESSENGER observed a well-defined flux transfer event (FTE) between 18:36:21 and 18:36:25 during its passage through the magnetosheath (Fig. 2). FTEs are produced by localized magnetic reconnection between the IMF and the planetary magnetic field at the MP (9). The magnetic field data in Fig. 3A show that this FTE was indeed preceded by a brief interval of southward IMF. Its flux rope topology is apparent, with the helical magnetic field surrounding and supporting the core region indicated by the bipolar B_y signature and the strong B_z, respectively. Given a typical anti-sunward magnetosheath flow speed of ~300 km s^{-1} and the ~4-s duration of the event, the size of this FTE is ~1200 km or ~0.5 R_M. Relative to Mercury's magnetosphere, this FTE is ~10 times larger than the size found at Earth (10). This result supports predictions that finite gyro-radius effects in Mercury's small magnetosphere will lead to relatively large FTEs (11).

When MESSENGER passed into Mercury's magnetotail (Fig. 2), there was a rapid transition to a quieter magnetic field directed predominantly northward but with a longitude angle near 0°, indicating that the spacecraft entered through the dusk flank of the tail into the central plasma sheet (12). The dominance of the B_z component over B_x and B_y components and the sunward longitude angle indicate that MESSENGER passed just north of the center of the cross-tail current sheet (Fig. 1). The high ratio of thermal to magnetic pressure typical of this region (12) is evident from the weakness of the magnetic field intensity in Mercury's tail at this point relative to the adjacent magnetosheath.

Between 18:47 and 18:49, the longitude angle of the magnetic field rotated from 0° (i.e., sunward) to near 180° (anti-sunward). This change indicates that MESSENGER moved southward through the cross-tail current sheet, consistent with its trajectory in Fig. 1. Around 19:00, the spacecraft altitude fell below ~800 km, and the magnetic field intensity began to increase quickly as MESSENGER moved into the region dominated by Mercury's dipolar planetary magnetic field (5). The increase in the magnetic field continued through closest approach and then decreased until MESSENGER exited the magnetosphere near the dawn terminator.

Examination of the high-resolution magnetic field longitude angle in Fig. 3B shows one 360° and several 180° rotations of the magnetic field in the X-Y plane between 18:43 and 18:46. The durations of the rotations ranged from ~10 to 25 s. Such rotations of the magnetic field in Earth's tail near the interface between the flanks of the plasma sheet and the magnetosheath are thought to be caused by vortices driven by the Kelvin-Helmholtz

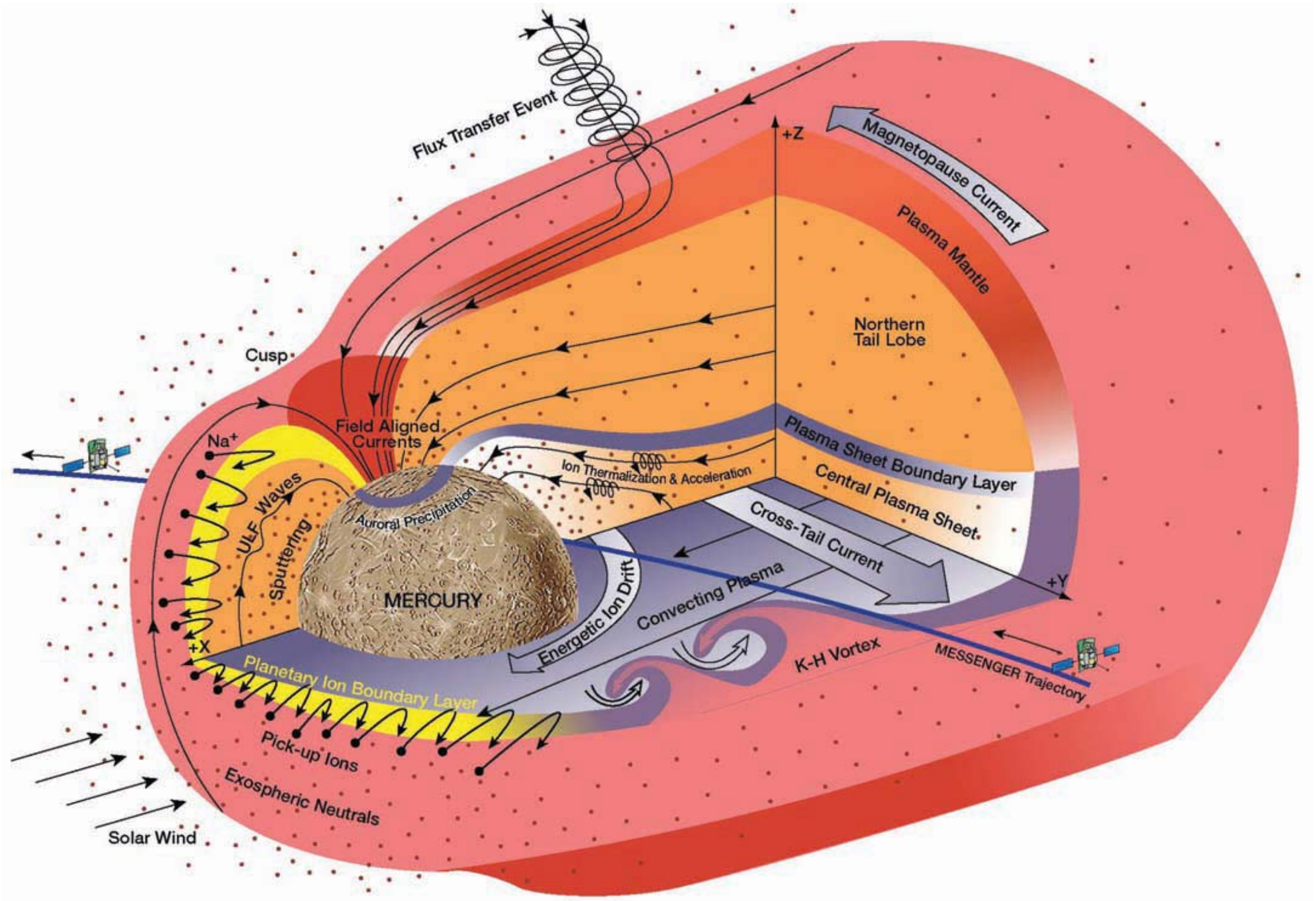


Fig. 1. Schematic of Mercury's magnetosphere highlighting the features and phenomena observed by MESSENGER, including the planetary ion boundary layer, large FTEs, flank K-H vortices, and ULF plasma waves.

(K-H) instability (*13*, *14*). Assuming Earth-like anti-sunward speeds of ~150 km s^{-1} (*14*) for these MESSENGER events, their implied spatial scale lengths are ~1 R_M. These scale lengths are smaller than similar features at Earth relative to the dimensions of their respective magnetospheres by a factor of ~3 (*14*).

The FIPS ion composition measurements show that Mercury's magnetosphere was permeated by planetary ions composed of Na^+ and other species in lesser amounts derived primarily from its exosphere (*15*). The coupling between these photoions and the magnetosphere has been the subject of extensive theory and modeling investigations (*16*–*20*) since sodium in Mercury's atmosphere was first detected telescopically from Earth (*21*). The spatial distribution of Na^+ (Fig. 2) represents a normalized count rate integrated over 3-min intervals (*7*). Further analysis is required to remove the effects of field-of-view obstructions and to determine bulk plasma properties such as density (*7*). The relative spatial distribution (Fig. 2) maximizes around closest approach, where the neutral atmosphere density should peak. This result is consistent with model predictions regarding the distribution of Na^+ within Mercury's magnetosphere (*17*). These models predict equatorial Na^+ densities along MESSENGER's near-tail trajectory that vary from 10^{-1} cm^{-3} to 10^{-2} cm^{-3} at dusk and dawn MPs, respectively (*17*). Secondary maxima in the FIPS Na^+ count rate exist just outside of the inbound and outbound MP crossings, indicating that the neutral sodium atmosphere extends to altitudes where photoions are strongly energized by pickup in the fast magnetosheath flow.

During the approach to Mercury, there were several intervals where the magnetic field decreased and its root mean square (RMS) variations increased (see horizontal bars in Fig. 2). Such variations are generally indicative of the growth of plasma waves stimulated by enhanced plasma density and/or temperature (*12*). The diamagnetic nature of these decreases is supported by the XRS count rates that increase around 19:00, coinciding with the first of these intervals (Fig. 2). The increase in XRS counts seen near 19:00 is believed to be due to fluorescence in the Mg- and Al-filtered

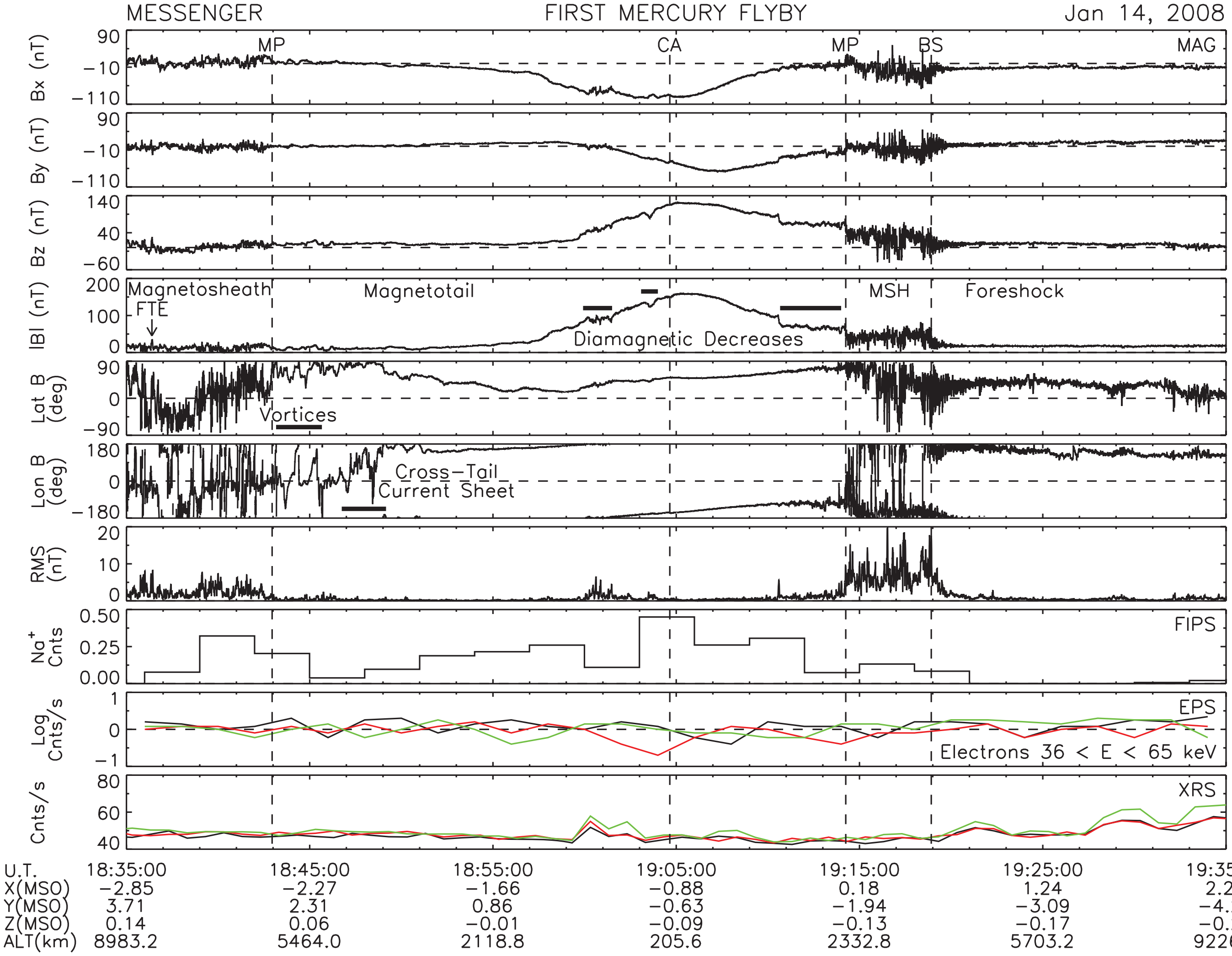


Fig. 2. Overview of MESSENGER magnetospheric measurements taken by the MAG, FIPS, EPS, and XRS instruments. Closest approach (CA) was at an altitude (ALT) of 201.4 km at 19:04:39 very near local midnight (00:04 local time). The magnetic field in Mercury solar orbital (MSO) coordinates is displayed in the top graphs along with the latitude and longitude direction angles and the RMS variance calculated over 3-s intervals. The MSO coordinate system is defined as X_{MSO} directed from the center of the planet toward the Sun; Z_{MSO}, normal to Mercury's orbital plane and positive toward the north celestial pole; and Y_{MSO}, positive in the direction opposite to orbital motion. The longitude angle of the magnetic field is defined to be 0° toward the Sun and increases counterclockwise looking down from the north celestial pole. The magnetic field latitude is +90° when directed northward and 0° when it is in the X_{MSO}-Y_{MSO} plane. U.T. designates universal time.

gas-proportional counters (GPCs) and to bremsstrahlung in both the Be window of the unfiltered GPC and the Be-Cu collimator in front of all three GPCs caused by electrons in the energy range ~1 to 10 keV. A similar response was seen in the GPCs on the Near-Earth Asteroid Rendezvous mission (*22*). The presence of enhanced fluxes of 1 to 10 keV electrons is consistent with these nightside diamagnetic decreases being due to the presence of hot plasma.

The strongest magnetic field decrease occurred after the narrow, MP-like current sheet encountered at 19:10:35. The orientation and thickness of this current sheet and the later MP current sheet are very similar, as can be inferred from the nearly identical variations in the magnetic field components (Fig. 4A). They differ primarily in intensity. The inner current sheet is only about half as strong as the MP current sheet. The difference in altitude between these two current sheets is ~1000 km. The enhanced RMS variations in the magnetic field indicate that the outer current sheet is the boundary between the magnetosphere and the magnetosheath and that the decreased magnetic field intensity between these two current sheets is due to enhanced plasma pressure [see also (*5*)].

This double MP signature had not been observed previously at Mercury or any other planetary magnetosphere. The decrease in the magnetic field in the outer part of the dawnside magnetosphere may be caused by the diamagnetic effect of solar wind plasma flowing into the magnetosphere along flux tubes opened by reconnection near the cusps or locally created planetary ions. Although magnetic reconnection is expected to be more effective in creating open flux at Mercury than at other planets (*23*), it has not been observed elsewhere to produce such broad boundary layers or multiple current sheets. Alternatively, the inner current sheet and the diamagnetic layer could be caused by hot planetary ions that enter the magnetosphere after being picked up and accelerated by the fast solar wind flow in the magnetosheath. At the dawn terminator, the magnetosheath flow speed would typically be ~300 km s^{-1}. For Na$^+$, the depth of penetration into the magnetosphere would be ~1 gyro-radius or ~1000 km, a value comparable to the observed thickness of the region of depressed magnetic field. If present in sufficient numbers, pickup ions entering the magnetosphere from the magnetosheath might create a planetary ion boundary layer bounded by an inner current sheet and the MP (Fig. 1).

The pickup process produces ion distributions that are unstable to the growth of ion cyclotron waves and other plasma-wave modes (*24*–*26*). No clear wave trains near the Na$^+$ cyclotron frequency are present in the MESSENGER measurements, consistent with Mariner 10 observations (*27*). During its closest approach and outbound passage, however, MESSENGER did observe ultralow-frequency (ULF) waves with frequencies of ~0.5 to 1.5 Hz, or just

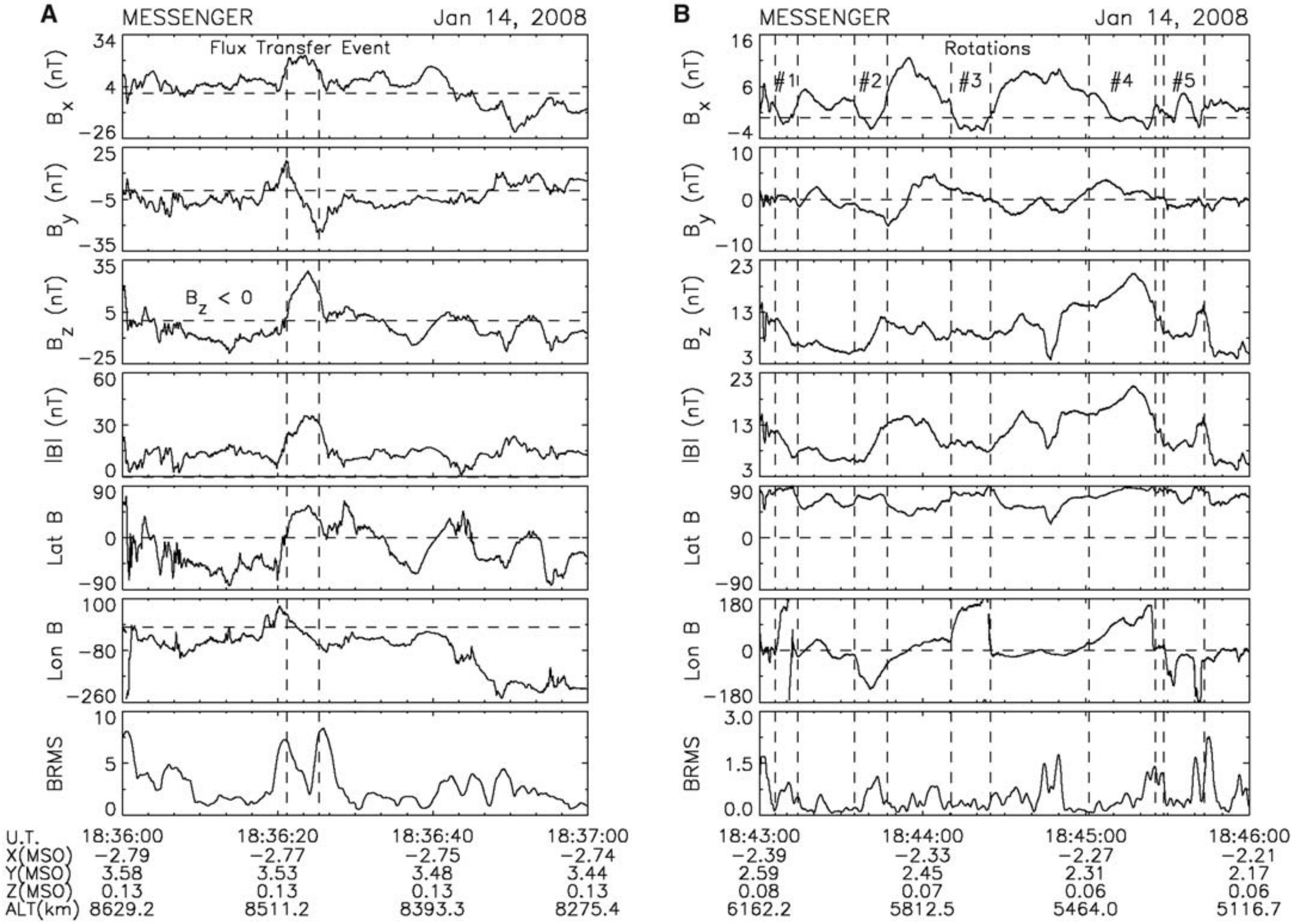


Fig. 3. (**A**) MESSENGER magnetic field observations of a large FTE in Mercury's magnetosheath. (**B**) Magnetic field observations of rotational signatures, possibly due to K-H–driven waves or vortices on the flanks of the magnetosphere.

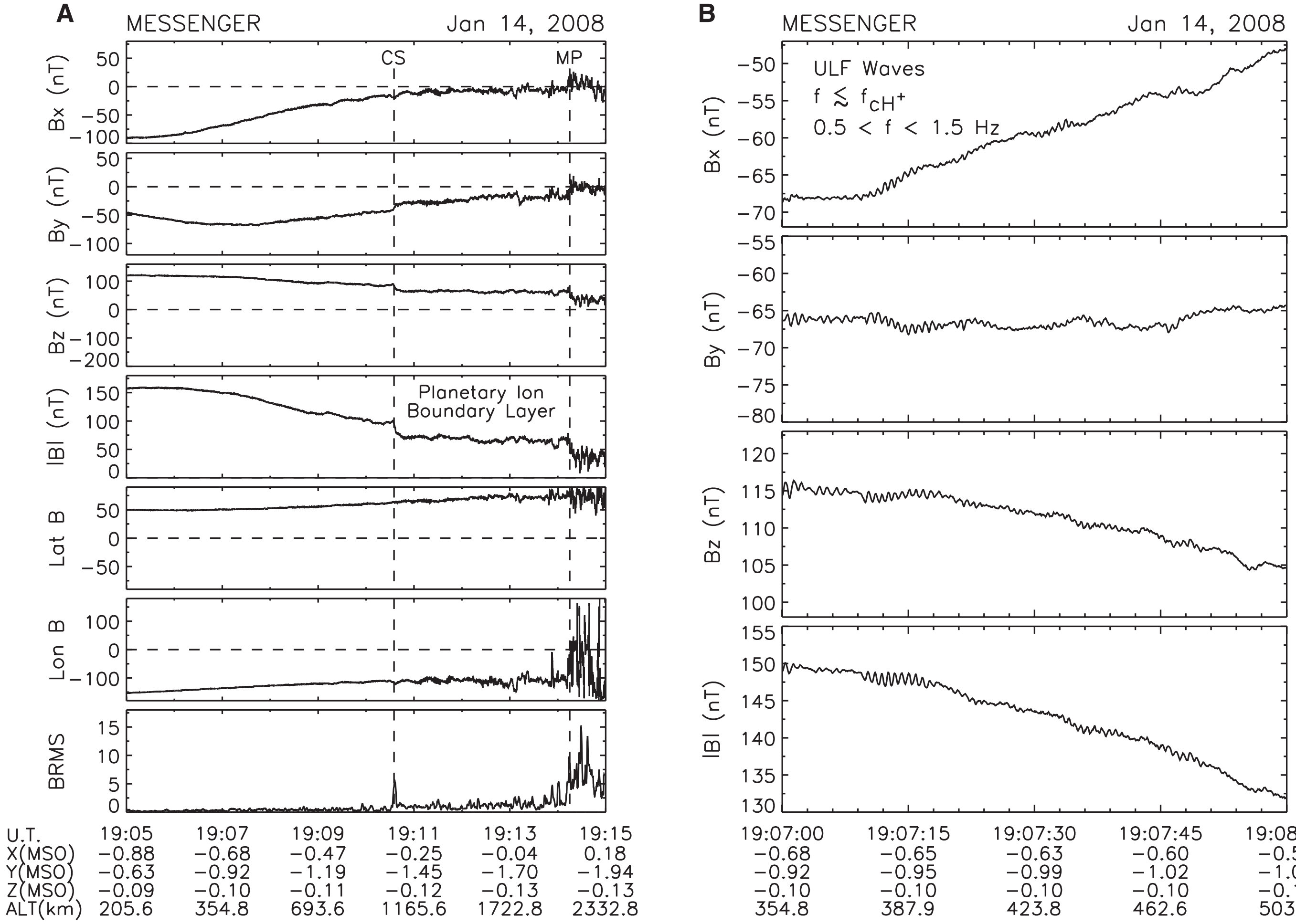


Fig. 4. (**A**) Magnetic field observations of the inner current sheet and MP boundary observed as MESSENGER exited the dawnside magnetosphere. (**B**) Magnetic field observations of ULF waves detected in Mercury's magnetosphere.

below the proton cyclotron frequency (f_{cH^+}) (Fig. 4B). They appear similar to the much shorter interval of ULF waves observed by Mariner 10 near closest approach during its first encounter (*28*). The frequency of these waves tended to increase with distance from Mercury until the outbound boundary layer was entered, where their frequency decreased and their amplitude increased to values as high as ~10 nT peak to peak.

MESSENGER has revealed Mercury's magnetosphere to be immersed in a cloud of cometlike planetary ions. Although the solar wind interaction appears dominated by Mercury's magnetic field, the presence of heavy planetary ions may exert influence from kinetic to magnetohydrodynamic scale lengths.

References and Notes

1. J. E. P. Connerney, N. F. Ness, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 494–513.
2. C. T. Russell, D. N. Baker, J. A. Slavin, in *Mercury*, F. Vilas, C. R. Chapman, M. S. Matthews, Eds. (Univ. of Arizona Press, Tucson, AZ, 1988), pp. 514–561.
3. S. C. Solomon *et al.*, *Planet. Space Sci.* **49**, 1445 (2001).
4. B. J. Anderson *et al.*, *Space Sci. Rev.* **131**, 417 (2007).
5. B. J. Anderson *et al.*, *Science* **321**, 82 (2008).
6. G. B. Andrews *et al.*, *Space Sci. Rev.* **131**, 523 (2007).
7. T. H. Zurbuchen *et al.*, *Science* **321**, 90 (2008).
8. C. E. Schlemm *et al.*, *Space Sci. Rev.* **131**, 393 (2007).
9. C. T. Russell, R. C. Elphic, *Space Sci. Rev.* **22**, 681 (1978).
10. C. T. Russell, R. J. Walker, *J. Geophys. Res.* **90**, 11067 (1985).
11. M. M. Kuznetsova, L. M. Zeleny, in *Proceedings of the Joint Varenna-Abastumani International School and Workshop on Astrophysics*, Sponsoring Organization, Sukhumi, USSR, 19 to 28 May 1986 (European Space Agency SP-251, Noordwijk, Netherlands, 1986), pp. 137–146.
12. J. A. Slavin *et al.*, *J. Geophys. Res.* **90**, 10875 (1985).
13. D. H. Fairfield *et al.*, *J. Geophys. Res.* **105**, 21159 (2000).
14. M. Fujimoto *et al.*, *J. Geophys. Res.* **103**, 4391 (1998).
15. W. E. McClintock *et al.*, *Science* **321**, 92 (2008).
16. W.-H. Ip, *Icarus* **71**, 441 (1987).
17. D. C. Delcourt *et al.*, *Ann. Geophys.* **21**, 1723 (2003).
18. K. Kabin, T. I. Gombosi, D. L. DeZeeuw, K. G. Powell, *Icarus* **143**, 397 (2000).
19. E. Kallio, P. Janhunen, *Ann. Geophys.* **21**, 2133 (2003).
20. P. Trávníček, P. Hellinger, D. Schriver, *Geophys. Res. Lett.* **34**, L05104, 10.1029/2006GL028518 (2007).
21. A. Potter, T. Morgan, *Science* **229**, 651 (1985).
22. R. D. Starr *et al.*, *Adv. Space Res.* **24**, 1159 (1999).
23. J. A. Slavin, R. E. Holzer, *J. Geophys. Res.* **84**, 2076 (1979).
24. A. L. Brinca, B. T. Tsurutani, *Astron. Astrophys.* **187**, 311 (1987).
25. K.-H. Glassmeier, D. Klimushkin, C. Othmer, P. Mager, *Adv. Space Res.* **33**, 1875 (2004).
26. L. G. Blomberg, J. A. Cumnock, K.-H. Glassmeier, R. A. Treumann, *Space Sci. Rev.* **132**, 575 (2007).
27. S. A. Boardsen, J. A. Slavin, *Geophys. Res. Lett.* **34**, L22106, 10.1029/2007GL031504 (2007).
28. C. T. Russell, *Geophys. Res. Lett.* **16**, 1253 (1989).
29. We thank all those who contributed to the success of the first MESSENGER flyby of Mercury. Science discussions with S. A. Boardsen and M. Sarantos and data visualization and graphics support by C. Liebrecht and M. Marosy are gratefully acknowledged. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington.

14 April 2008; accepted 3 June 2008
10.1126/science.1159040

REPORT

MESSENGER Observations of the Composition of Mercury's Ionized Exosphere and Plasma Environment

Thomas H. Zurbuchen,[1]* Jim M. Raines,[1] George Gloeckler,[1] Stamatios M. Krimigis,[2,3] James A. Slavin,[4] Patrick L. Koehn,[5] Rosemary M. Killen,[6] Ann L. Sprague,[7] Ralph L. McNutt Jr.,[2] Sean C. Solomon[8]

The region around Mercury is filled with ions that originate from interactions of the solar wind with Mercury's space environment and through ionization of its exosphere. The MESSENGER spacecraft's observations of Mercury's ionized exosphere during its first flyby yielded Na^+, O^+, and K^+ abundances, consistent with expectations from observations of neutral species. There are increases in ions at a mass per charge (m/q) = 32 to 35, which we interpret to be S^+ and H_2S^+, with $(S^+ + H_2S^+)/(Na^+ + Mg^+) = 0.67 \pm 0.06$, and from water-group ions around m/q = 18, at an abundance of 0.20 ± 0.03 relative to Na^+ plus Mg^+. The fluxes of Na^+, O^+, and heavier ions are largest near the planet, but these Mercury-derived ions fill the magnetosphere. Doubly ionized ions originating from Mercury imply that electrons with energies less than 1 kiloelectron volt are substantially energized in Mercury's magnetosphere.

Since the discovery of Mercury's internal magnetic field during the Mariner 10 flyby encounters in 1974–1975 (*1*), there has been speculation about the nature of the interactions of the solar wind and electromagnetic fields with Mercury's surface and near-planetary exosphere (*2*). These interactions were surmised from remote measurements of the neutral exosphere (*3*) because Mariner 10 did not provide any direct observations of the ionized component or characterize Mercury's ion plasma environment. Because of the comparatively small size of Mercury's magnetic field, its tenuous atmosphere, and its close proximity to the Sun, Mercury's surface is subject to more direct space-weathering interactions than are those of other terrestrial planets. In addition to ejecting neutral particles that form Mercury's exosphere, surface-sputtering interactions can lead directly to ionized components. Such newly formed ions can also originate in the exosphere through ionization. In either case, these ions are created approximately at rest near the planet and then undergo energization by electromagnetic processes that dominate Mercury's space environment. The MESSENGER spacecraft measured a mass-per-charge (m/q) spectrum of ions in Mercury's exosphere during its first flyby on 14 January 2008 (Fig. 1). These measurements were performed by the Fast Imaging Plasma Spectrometer (FIPS), the low-energy portion of the Energetic Particle and Plasma Spectrometer (EPPS) instrument (*4*) aboard MESSENGER.

Here, we focus on the relative abundances of ions in the m/q range between 4 and ~50 atomic mass units (amu) per unit of charge (e). Ions with a m/q <4 (H^+ and He^{++}) largely originate from the solar wind; ions with a m/q >10 are generally produced locally. Ions with a m/q of 23 to 24 (Na^+ plus Mg^+) are clearly the most dominant heavy ions (Table 1). Neutral Na has been observed remotely from Earth (*3*) and also during the MESSENGER flyby (*5*).

Although Na^+ dominates, several secondary peaks (around m/q = 16 to 18, 32 to 36, 28, and 39 to 40) also stand out. We identified these peaks, respectively, as predominantly O^+ and water-group ionized molecules; S^+ and H_2S^+; and the surface-bound mineral components Si^+, K^+, and Ca^+. We cannot rule out additional contributions to ions in the dominant peaks from nearby elements and various molecular species (some are listed in Table 1).

The abundances of Si and especially of Na and S relative to O in the solar wind (*6*) are too low and their ionization states too high to account for the abundances of these ions. Their source is, therefore, either Mercury's surface or its exosphere.

When inbound, MESSENGER passed through the plasma sheet, the region between the two lobes of the magnetotail. This region is a natural magnetospheric reservoir for hot plasma with energies up to at least several thousand electron volts and densities of at least 1 cm^{-3} (*7*). Plasma-sheet electrons are most likely to be the ionization source for the creation of multiply charged ions (such as O^{++}) observed by FIPS from corresponding singly ionized atoms (such as O^+). MESSENGER does not directly measure thermal and suprathermal electrons. However, our detec-

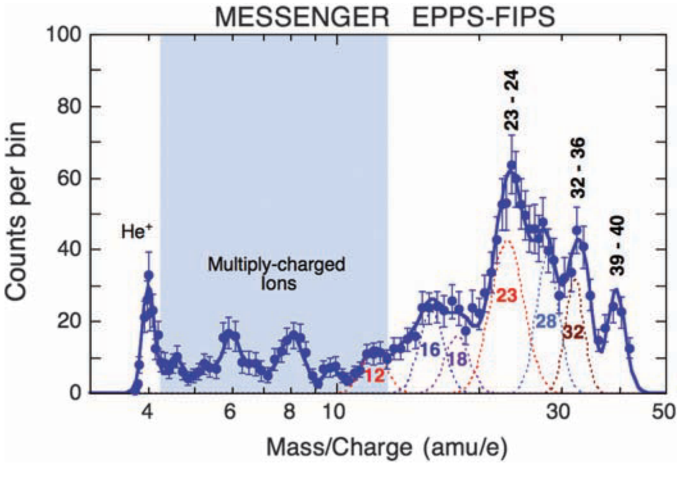


Fig. 1. Counts from the FIPS sensor per m/q bin versus m/q of ions with $3.8 < m/q < 42$ accumulated in Mercury's magnetosphere between 18:43 and 19:14 UTC during the flyby on 14 January 2008. Thin curves are Gaussian fits to several major peaks of the m/q histogram, and the thick curve is the sum of all Gaussian distributions. Multiply charged ions are observed primarily below $m/q \approx 12$. FIPS measures the energy per charge (E/q) on an ion from 0.1 to 13.5 keV/e, its arrival direction with an angular resolution of ~15°, and the m/q ratio (derived from E/q and a time-of-flight measurement) determined to an accuracy $\sigma(m/q) = \Delta(m/q)/(m/q)$ that ranges from 0.04 to 0.08, depending on the mass of the ion. Because of limited counting statistics, we followed a minimum-least-squares procedure to estimate the relative abundance of an ion at a given allowed m/q using log-Gaussian distributions with $\sigma(m/q)$ calculated from preflight calibrations.

[1] Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, 2455 Hayward Street, Ann Arbor, MI 48109–2143, USA. [2] Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA. [3] Academy of Athens, Research Center for Astronomy and Applied Mathematics, Soranou Efessiou 4, Athens 11527, Greece. [4] Heliophysics Science Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [5] Department of Physics and Astronomy, Eastern Michigan University, 303 Strong Hall, Ypsilanti, MI 48197, USA. [6] Department of Astronomy, University of Maryland, College Park, MD 20742, USA. [7] Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [8] Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA.

*To whom correspondence should be addressed. E-mail: thomasz@umich.edu

tion of multiply charged ions provides indirect evidence for the presence of a hot electron component near the planet and supports the indirect detection of ~1- to 10-keV electrons in this region by MESSENGER's x-ray spectrometer (*8*, *9*).

The variation of the relative count rates of the key components of Mercury's plasma environment as sampled by MESSENGER can be correlated with the global structure of Mercury's magnetosphere (*10*, *11*) (Fig. 2), including two crossings of the magnetospheric bow shock and two crossings of the magnetopause (which separates solar wind from magnetospheric plasma). The maximum intensity of the heaviest ions occurred at the point of closest approach, 201.4 km above Mercury's surface (Fig. 2, top).

The energy distribution of protons originating from the solar wind substantially increased at the bow shock (Fig. 2A). Unperturbed solar wind was not directly measured by FIPS because of the sensor's location behind MESSENGER's sunshade. Within the magnetosphere, however, protons are easily measured because of a decrease in Mach number at the bow shock and deflections of the solar wind from the solar direction. Two distinct flux enhancements from 18:10 to 18:30 UTC were caused by spacecraft rotations about the solar direction. He^{++} ions, a direct indicator of solar-wind plasma (Fig. 2B), were found throughout the magnetospheric pass. The flux in the m/q range $3 < m/q < 10$ (Fig. 2C) shows behavior in qualitative agreement with the spatial distribution of He^{++} ions, indicating in turn either a solar-wind source or a direct coupling of the ions to such a solar-wind source through a charge exchange of Mercury-derived, singly charged ions with the solar wind. This connection includes multiply charged ions, such as O^{++} and C^{++}. A clear dominance of these contributions was found in the vicinity of the magnetopause, marked with blue dashed lines in Fig. 2.

The distribution of ions with $10 < m/q < 42$ is dominated by peaks at m/q = 16 to 18 and m/q = 23 to 24 (Fig. 2, D and E). The large number of these ions implies that the ionized exosphere of Mercury extends throughout the magnetosphere and therefore must have an extended source.

For the heavy ions observed by FIPS that correspond to previously known neutral species Na, K, and Ca, their relative counts are consistent with photoionization, given known exospheric column abundances and photoionization rates (*12*). Because most of the heavy ions were seen near the point of closest approach, where the spacecraft traversed the plasma sheet, electron-impact ionization of neutral species probably also contributes to the observed ion population. The apparent dawn-dusk asymmetry may be attributed to the flyby geometry: The spacecraft entered the magnetopause far down the dusk flank and exited the magnetosphere on the near-dawn dayside, where the plasma density is expected to be greater (Fig. 2, top). Some of the asymmetry may further be caused by ion gyroradius effects. The peaks at m/q = 32 and 28 are consistent with O_2^+ and Si^+, respectively, which can be produced by dissociative ionization of silica. Another peak at m/q = 32 may be S^+, as earlier predicted (*13*). The Mg^+ (m/q = 24) in Mercury's exosphere is consistent with the identification of Mg-rich pyroxene in Mercury surface materials (*14*).

The spatial distribution of species within the ionized exosphere reflects the large-scale structure of Mercury's magnetosphere. Heavy

Table 1. Abundance ratios of possible ions and molecular ions relative to Na^+ plus Mg^+.

Mass/charge (amu/e)	Representative ion or molecular ion species	Abundance ratio*
23–24	Na^+, Mg^+	1.00
32–35	S^+, O_2^+, H_2S^+	0.67 ± 0.06
28	Si^+, Fe^{++}	0.53 ± 0.06
39–40	K^+, Ca^+	0.44 ± 0.05
17–19	H_2O^+, H_3O^+, OH^+	0.20 ± 0.03
4.67–11	Multiply charged ions	0.20 ± 0.03
16	O^+	0.20 ± 0.03
14	N^+, Si^{++}	0.09 ± 0.02
11–12	C^+, Na^{++}, Mg^{++}	0.08 ± 0.02
4	He^+	0.03 ± 0.01

*Abundance (relative to the sum of Na^+ and Mg^+) of the sum of ions and molecular ions listed in column 2. Uncertainties are dominated by limited counting statistics.

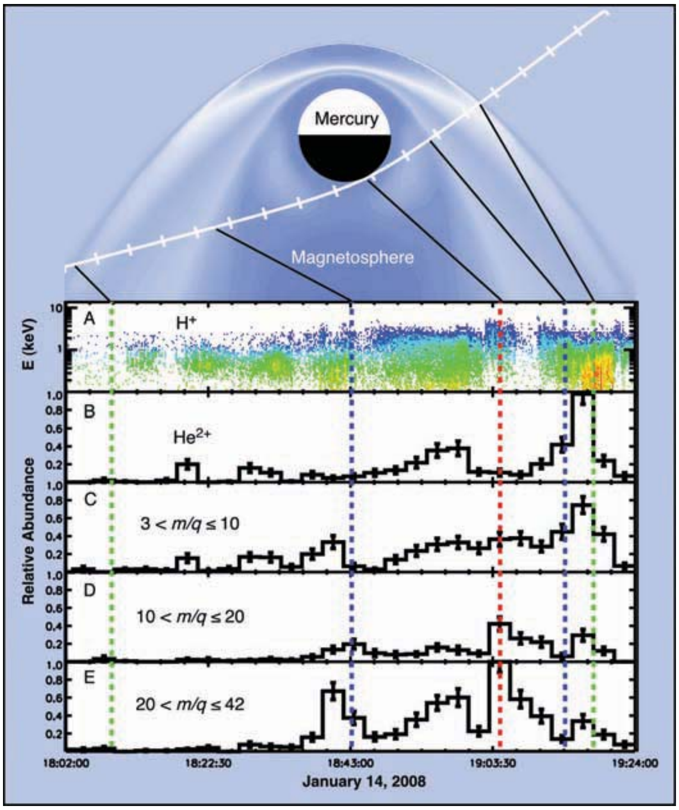


Fig. 2. Spatial distribution of key plasma components in relation to magnetospheric structure (*11*). (Top) Overview of magnetospheric geometry from a magnetohydrodynamic simulation (*10*) used to order the timeline of the MESSENGER flyby and the locations and encounter times of key components of the Mercury space environment. (**A** and **B**) Energy distribution of protons (at a resolution of 8 s, during which FIPS performs a complete *E*/*q* stepping sequence from 0.1 to 13 keV/*e*) and normalized He^{++} flux. Both species originate in the solar wind. The temporal variability of these components is associated with changes in plasma characteristics, as well as with temporal variability of the obstruction geometry, especially for solar wind protons. (**C**, **D**, and **E**) Normalized fluxes of ions in specified m/q ranges. All fluxes [(B) to (E)] are normalized to the peak flux in He^{++}. The fluxes of heavy ions with $10 < m/q < 42$ maximize near the planet but are also found throughout the magnetosphere. During the flyby, the spacecraft sunshade, one of the solar panels, and other spacecraft structures limit the field of view of FIPS to ~π steradians. Vertical dashed lines denote the crossing of the bow shock (green), the magnetopause (blue), and the point of closest approach (red), based on magnetic field data (*9*, *11*).

ions (masses between 10 and 50 amu) are most abundant near Mercury's surface, between local midnight and dawn (Fig. 2). When ionized, they are quickly accelerated by electric fields expected near the planet toward the nightside surface (*15*). This process contributes to the recycling of exospheric constituents and to the dawn enhancement of Na and K in the exosphere and inhibits the loss of material to the interplanetary medium (*16*). Molecular ion species, such as H_2O^+ and H_2S^+ observed by FIPS, are probably associated with chemical sputtering of the surface (*17*) or sputtering of cold-trapped ices.

References and Notes

1. N. F. Ness, K. W. Behannon, R. P. Lepping, Y. C. Whang, K. H. Schatten, *Science* **185**, 151 (1974).
2. J. A. Slavin *et al.*, *Space Sci. Rev.* **131**, 133 (2007).
3. A. E. Potter, T. H. Morgan, *Science* **248**, 835 (1990).
4. G. B. Andrews *et al.*, *Space Sci. Rev.* **131**, 523 (2007).
5. W. E. McClintock *et al.*, **321**, 92 *Science* (2008).
6. R. von Steiger *et al.*, *J. Geophys. Res.* **105**, 27217 (2000).
7. T. Mukai, K. Ogasawara, Y. Saito, *Adv. Space Res.* **33**, 2166 (2004).
8. P. Mazzotta, G. Mazzitelli, S. Colafrancesco, N. Vittorio, *Astron. Astrophys.* **133** (Supplement), 403 (1998).
9. J. A. Slavin *et al.*, *Science* **321**, 85 (2008).
10. K. Kabin, T. I. Gombosi, D. L. DeZeeuw, K. G. Powell, *Icarus* **143**, 397 (2000).
11. B. J. Anderson *et al.*, *Science* **321**, 82 (2008).
12. W. F. Huebner, J. J. Keady, S. P. Lyon, *Solar Photo Rates for Planetary Atmospheres and Atmospheric Pollutants* (Kluwer Academic, Dordrecht, Boston, 1992).
13. A. L. Sprague, D. M. Hunten, K. Lodders, *Icarus* **118**, 211 (1995).
14. A. L. Sprague, T. L. Roush, *Icarus* **133**, 174 (1998).
15. D. C. Delcourt *et al.*, *Ann. Geophys.* **21**, 1723 (2003).
16. A. L. Sprague, *J. Geophys. Res.* **97**, 18257 (1992).
17. A. E. Potter, *Geophys. Res. Lett.* **22**, 3289 (1995).

18. The MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington.

18 April 2008; accepted 30 May 2008
10.1126/science.1159314

REPORT

Mercury's Exosphere: Observations During MESSENGER's First Mercury Flyby

William E. McClintock,[1]* E. Todd Bradley,[2] Ronald J. Vervack Jr.,[3] Rosemary M. Killen,[4] Ann L. Sprague,[5] Noam R. Izenberg,[3] Sean C. Solomon[6]

During MESSENGER's first Mercury flyby, the Mercury Atmospheric and Surface Composition Spectrometer measured Mercury's exospheric emissions, including those from the antisunward sodium tail, calcium and sodium close to the planet, and hydrogen at high altitudes on the dayside. Spatial variations indicate that multiple source and loss processes generate and maintain the exosphere. Energetic processes connected to the solar wind and magnetospheric interaction with the planet likely played an important role in determining the distributions of exospheric species during the flyby.

Every solid object in the solar system possesses an interface that separates its surface from the external space environment. Mercury's interface is a surface-bounded exosphere in which the constituent atoms and molecules travel on nearly ballistic trajectories and are far more likely to collide with the surface than with each other. Mercury's exosphere is known to contain hydrogen (H), helium (He), and oxygen (O), which were detected by the Mariner 10 Ultraviolet Spectrometer (*1*, *2*), and sodium (Na), potassium (K), and calcium (Ca), which were discovered with ground-based telescopes (*3*–*5*). Although other airless bodies in the solar system also have a surface-bounded exosphere, Mercury's is unique for several reasons. The planet's highly elliptical orbit causes large seasonal changes in exosphere properties (*6*, *7*). The interaction of Mercury's planetary magnetic field with the interplanetary medium modulates the spatial distribution and flux of solar wind plasma and high-energy charged particles that sputter materials from the surface to populate the exosphere. This modulation results in large density variations on time scales as short as a few hours (*8*). Neutral species released from the surface with sufficient energy are accelerated by solar radiation pressure to form an extended antisunward tail of atoms, as demonstrated by ground-based observations of Mercury's neutral Na tail (*9*–*11*). Here, we report on observations of Mercury's exosphere and Na tail

[1]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA. [2]Department of Physics, University of Central Florida, Orlando, FL 32816, USA. [3]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723, USA. [4]Department of Astronomy, University of Maryland, College Park, MD 20742, USA. [5]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA. [6]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA.

*To whom correspondence should be addressed. E-mail: william.mcclintock@lasp.colorado.edu

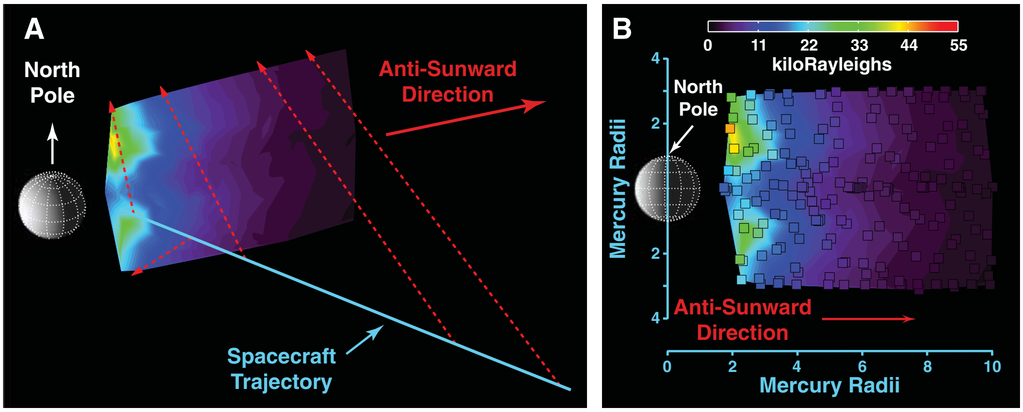


Fig. 1. Viewing geometry and observed D2 radiances during the Na tail observations. (**A**) Representative line-of-sight vectors (dashed red lines) along the inbound trajectory (blue line) from the spacecraft to the plane defined by Mercury's north pole and the Sun-Mercury line. (**B**) Projection of the observed line-of-sight radiances into the defined plane [perspective view is also shown in (A)]. This projection is an interpolated contour image based on the actual observations, which are overplotted on the image as squares color-coded to the observed radiances. The match between the observations and the interpolated image is good, indicating that the interpolation did not introduce artifacts. The clustering of observations near the Sun-Mercury line at distances of 4.5 and 5.7 Mercury radii reflect times when pointing was altered during acquisition of surface images.

obtained with the Ultraviolet and Visible Spectrometer (UVVS) channel of the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) (*12*) during the 14 January 2008 flyby of Mercury by the MESSENGER spacecraft (*13*).

The MASCS instrument consists of a compact Cassegrain telescope that simultaneously feeds the UVVS channel and a Visible and Infrared Spectrograph (VIRS) channel. The UVVS is a scanning grating monochromator that covers the wavelength range 115 to 610 nm, with spectral resolution of ~0.5 nm, and was designed to measure resonantly scattered sunlight from known exospheric constituents (H, O, Na, K, Ca), as well as predicted species such as magnesium (Mg), iron (Fe), aluminum (Al), and sulfur (S) (*14*, *15*).

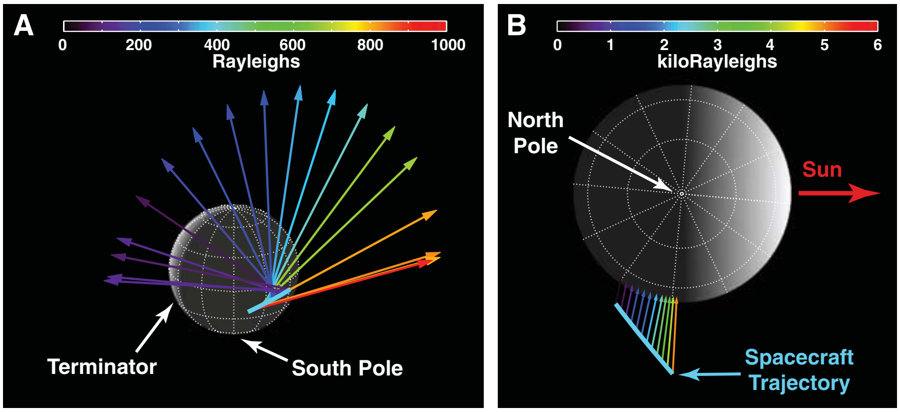


Fig. 2. Viewing geometry and observed radiances during the Ca near-tail (**A**) and near-terminator (**B**) observations. (A) UVVS line-of-sight vectors color-coded according to the observed radiances. (B) Similar vectors as the spacecraft emerges from Mercury's shadow. The observed radiances here correspond to columns of illuminated Ca atoms between MESSENGER and Mercury's surface.

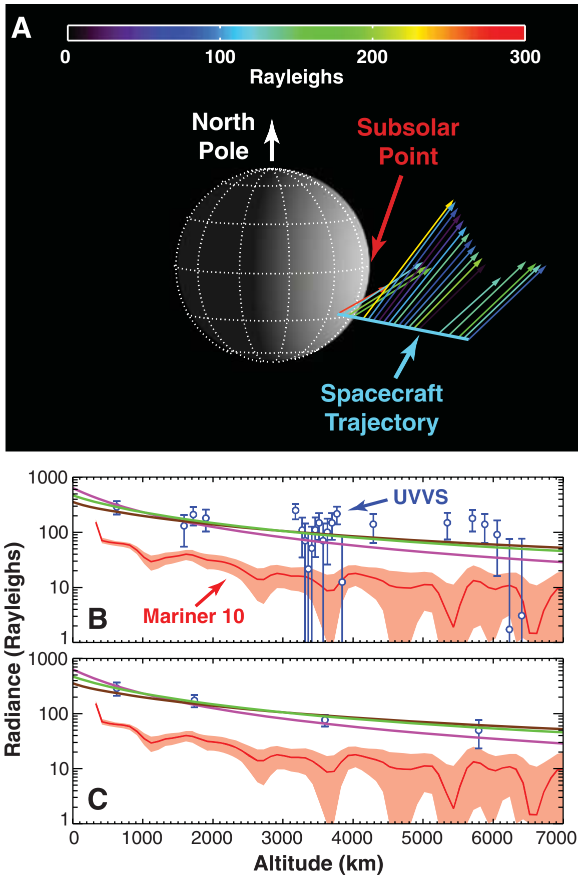


Fig. 3. Viewing geometry and observed radiances during the H dayside observations and comparison with Mariner 10 measurements. (**A**) UVVS line-of-sight vectors color-coded according to the observed radiances. The observations move in a generally radial direction owing to spacecraft motion along the trajectory, whereas motion perpendicular to Mercury's equatorial plane is caused by the spacecraft rolling to enable acquisition of images of Mercury's surface. Vector lengths are a function of the projection of three-dimensional vectors onto a two-dimensional image and convey no information about the brightness; only the color is indicative of the emission strength. (**B**) Comparison of the UVVS H altitude profile (blue symbols) to the profile (red line; shaded uncertainty range) measured by Mariner 10 during its first encounter with Mercury (*17*) as well as to three model curves (see text): T = 250 K, n_0 = 140 cm^{-3} (magenta); T = 420 K, n_0 = 90 cm^{-3} (green); T = 800 K, n_0 = 70 cm^{-3} (brown); T is the exospheric temperature and n_0 is the exobase density. (**C**) Same as (B) but with the UVVS data binned into altitude ranges (0 to 1000 km, 1000 to 2500 km, 2500 to 4500 km, and >4500 km) to illustrate the decrease in radiance with altitude more clearly. Although the UVVS observations indicate stronger overall emission than during the Mariner 10 encounter, the altitude-binned UVVS profile shows that the scale heights are consistent with one another.

Observations of Mercury's exosphere were obtained in three distinct regions: tail, near-tail/near-terminator, and dayside. Details of the flyby geometry are described in (*16*). In the tail region, the UVVS was configured to observe emission from the H Lyman α line at 121.6 nm and the D lines of Na at 589.0 nm (D2) and 589.6 nm (D1). The tail observations were obtained by scanning the UVVS 0.1° × 1.0° field of view up and down across the plane defined by the Sun-Mercury line and Mercury's north pole (Fig. 1). These observations started down the tail ~24,500 km from Mercury along the antisunward direction and continued to a radial distance of ~4500 km as the spacecraft was inbound to the planet. The observations were designed to sweep the line-of-sight intercept with the plane through a fixed distance of ±3 Mercury radii (R_M, 1 R_M = 2440 km) about the Sun-Mercury line. However, the angle relative to the Sun-Mercury line required to reach the ±3 R_M limit increased monotonically as the spacecraft approached the planet (Fig. 1A). Therefore, the line of sight swept through a larger angular region of space as the tail observations progressed, viewing emission along lines of sight that sampled larger ranges of latitude with time.

Strong Na emission was observed throughout the entire tail region (Fig. 1). The most striking features were that emissions were stronger in the north than in the south and were relatively weak in the near-equatorial regions. Observing geometry caused the UVVS line of sight to sweep across Mercury's shadow, complicating density determinations because the UVVS observes resonance lines that are excited by solar illumination (i.e., atoms may exist in the shadow region but do not emit). The observed relative strength of the emission as a function of look direction, however, is a robust measurement because the UVVS effectively integrates from the spacecraft to infinity. In contrast to Na, after accounting for the changing viewing geometry, no significant H emission was observed in the tail region above the background interplanetary H Lyman α emission.

As the spacecraft entered Mercury's shadow, it executed a 180° roll in order to position the remote sensing instruments to observe the sunlit surface after emergence from shadow. Although the spacecraft remained in shadow during the entire roll, the UVVS line of sight extended out from the shadow through illuminated regions of the exosphere. During this traverse, the UVVS was configured to observe Ca at 422.7 nm and K at 404.6 nm in the near-tail exosphere (close to Mercury on the tail side of the planet). The observations started with the line of sight pointed toward the dawn hemisphere near the equator, then swept from the equator up to the north pole and back down to the equator on the dusk side of the planet (Fig. 2A). At the beginning of the roll, nearly simultaneous observations of Na and Ca in the direction of the dawn side of the planet yielded a Na/Ca ratio of 10 ± 0.2. Unexpectedly, Ca emissions were stronger in the dawn hemisphere and decreased as the line of sight rolled toward dusk.

Ca observations continued as the spacecraft emerged from shadow and the line of sight intersected the surface on the nightside of the planet (Fig. 2B). These observations sampled the illuminated column between the spacecraft and Mercury's surface, in contrast to all other observations during the flyby, for which the line of sight remained completely off the surface. Ca observations continued until the line of sight crossed the terminator into the sunlit hemisphere, at which point surface observations began. At the end of these near-terminator observations, additional Na measurements were also obtained, indicating that the Na/Ca ratio was 50 ± 6 along the 1490-km-long path from the spacecraft to the surface near the dawn terminator. Potassium was not detected in observations in the same near-terminator region (K/Na < 0.07), consistent with a less dense K exosphere and a higher photoionization rate (the principal loss mechanism) for K than for Na.

The final region observed during the flyby was the dayside exosphere. These observations were obtained as the UVVS line of sight drifted off the planetary limb and covered the region from approximately the surface up to 6500 km (Fig. 3A). For spacecraft operational reasons, the UVVS line of sight did not drift at a uniform rate radially but rather moved approximately radially above the subsolar point (up to 15° off the Sun-Mercury line) with a varying rate that resulted in a nonconstant altitude sampling. The measured H radiances (corrected for background interplanetary H emission) are scattered but are generally higher than those seen by Mariner 10 during its first encounter (*17*) (Fig. 3B). The relatively large scatter in the UVVS data may be partly related to the changing geometry of the observations but may also be an indication of small-scale variations in the H density. The Mariner 10 data were observed with a field of view that spanned the entire dayside hemisphere from north to south and were therefore averaged over large spatial distances. The UVVS field of view is substantially smaller and as such is more sensitive to potential local variations in density.

Dayside Na observations suffered from scattered light from the bright sunlit surface. One dayside Na measurement was extracted and showed a Na D2 radiance of 269,000 Rayleighs at an altitude of 582 km above the subsolar point, consistent with a hot Na corona (*18*).

The Na tail is composed of the escaping component of the neutral Na exosphere, accelerated to escape velocity (4.2 km/s) or greater by solar radiation pressure. The observed line-of-sight column abundance of 1.7×10^9 atoms cm^{-2} at two Mercury radii down the tail, representing half a traverse across the entire tail, implies a lower limit on the escape rate of ~6×10^{23} atoms s^{-1} for an assumed initial antisunward velocity of 2 km/s. Photon-stimulated desorption (PSD) is expected to be a major source of Na to the exosphere [e.g., (*19*, *20*)], and we estimate a total PSD yield of 10^7 cm^{-2} by assuming that (i) the Na surface concentration is 0.5%, (ii) the PSD rate is one-third of the laboratory rate to account for regolith trapping (*21*), and (iii) the yield over the dayside hemisphere is variable to account for ion-enhanced rates at high latitudes (*22*). If we further assume that the ion-sputter and impact-vaporization yields are both equal to 20% of the PSD yield (*23*), then the total flux of Na from the surface is 5×10^{24} atoms s^{-1}. Under these assumptions, the escape rate is on the order of 10% of the total ejection from the surface.

The Na/Ca ratios are consistent with a hot Ca corona as discovered by ground-based measurements (*5*, *24*). Near the terminator, the observed Na distribution is likely a result of both PSD and ion-sputtered components. The source of Ca is less certain, but meteoroid volatilization and sputtering are likely candidates (*5*). Farther from the planet, in the near-tail region, the smaller ratio probably reflects a smaller scale height for the bulk of the Na exosphere than that for Ca.

The dayside H limb scan is consistent with the 420 K exospheric temperature measured during the Mariner 10 flyby (*17*); however, insufficient data were obtained to constrain the temperature rigorously. Models generated using the formalism of Chamberlain (*25*) and spanning 250 to 800 K with exobase densities of 140 to 70 cm^{-3} are also plausible within the uncertainties (Fig. 3, B and C). The overall higher H radiances observed by MASCS relative to the Mariner 10 data may have several causes, including differences in instrument calibration, incident solar flux, and exospheric H density.

Local enhancements and asymmetries in the distribution of exospheric species at Mercury have previously been ascribed to solar wind ion sputtering (*7*, *23*) or surface compositional and physical differences (*26*, *27*). The potential enhancement of the Na tail emission toward the northern hemisphere (Fig. 1) is consistent with a more energetic source of some exospheric Na atoms and suggests a solar wind–magnetosphere influence on the release mechanisms. Ca emission was observed to be strong in the dawn hemisphere and weak on the dusk side (Fig. 2), although whether the Ca release mechanism is related to the solar wind is yet to be determined. Hybrid simulations of the magnetosphere under conditions similar to those during the MESSENGER flyby (*28*) suggest both a strong asymmetry in the cusp plasma densities and a dawn-dusk asymmetry in plasma close to the planet, with higher densities at dawn, consistent with the Na high-latitude enhancement and the Ca dawn-dusk asymmetry in the MASCS observations. The combination of a spatially varying neutral exosphere, extended tail, and weak magnetic field yields a planetary environment like no other in the solar system (*29*).

References and Notes

1. A. L. Broadfoot, S. Kumar, M. J. S. Belton, M. B. McElroy, *Science* **185**, 166 (1974).
2. D. E. Shemansky, A. L. Broadfoot, *Rev. Geophys. Space Phys.* **15**, 491 (1977).
3. A. Potter, T. Morgan, *Science* **229**, 651 (1985).
4. A. E. Potter, T. H. Morgan, *Icarus* **67**, 336 (1986).
5. T. Bida, R. M. Killen, T. H. Morgan, *Nature* **404**, 159 (2000).
6. F. Leblanc *et al.*, *Icarus* **185**, 395 (2006).
7. A. E. Potter, R. M. Killen, T. H. Morgan, *Icarus* **186**, 571 (2007).
8. R. M. Killen, M. Sarantos, A. E. Potter, P. Reiff, *Icarus* **171**, 1 (2004).
9. A. E. Potter, R. M. Killen, T. H. Morgan, *Meteorit. Planet. Sci.* **37**, 1165 (2002).
10. A. E. Potter, R. M. Killen, *Icarus* **194**, 1 (2008).
11. J. Baumgardner, J. Wilson, M. Mendillo, *Geophys. Res. Lett.* **35**, L03201 (2008).
12. W. E. McClintock, M. R. Lankton, *Space Sci. Rev.* **131**, 481 (2007).
13. S. C. Solomon *et al.*, *Space Sci. Rev.* **131**, 3 (2007).
14. T. H. Morgan, R. M. Killen, *Planet. Space Sci.* **45**, 81 (1997).
15. A. L. Sprague, D. M. Hunten, K. Lodders, *Icarus* **118**, 211 (1995).
16. S. C. Solomon *et al.*, *Science* **321**, 59 (2008).
17. A. L. Broadfoot, D. E. Shemansky, S. Kumar, *Geophys. Res. Lett.* **3**, 577 (1976).
18. R. M. Killen, A. E. Potter, A. Fitzsimmons, T. H. Morgan, *Planet. Space Sci.* **47**, 1449 (1999).
19. B. V. Yakshinskiy, T. E. Madey, *Nature* **400**, 642 (1999).
20. R. M. Killen *et al.*, *Space Sci. Rev.* **132**, 433 (2007).
21. T. A. Cassidy, R. E. Johnson, *Icarus* **176**, 499 (2005).
22. M. Sarantos, R. M. Killen, A. S. Sharma, J. A. Slavin, *Geophys. Res. Lett.* **35**, L04105 (2008).
23. R. M. Killen *et al.*, *J. Geophys. Res.* **106**, 20,509 (2001).
24. R. M. Killen, T. Bida, T. H. Morgan, *Icarus* **173**, 300 (2005).
25. J. W. Chamberlain, *Planet. Space Sci.* **11**, 901 (1963).
26. A. L. Sprague, W. J. Schmitt, R. E. Hill, *Icarus* **136**, 60 (1998).
27. A. L. Sprague *et al.*, *Icarus* **129**, 506 (1997).
28. P. Trávníček, P. Hellinger, D. Schriver, *Geophys. Res. Lett.* **34**, L05104 (2007).
29. J. A. Slavin *et al.*, *Science* **321**, 85 (2008).

30. We thank M. Lankton, M. Kochte, and N. Mouawad for help in acquiring and analyzing these data. The MESSENGER project is supported by the NASA Discovery Program under contract NAS5-97271 to the Johns Hopkins University Applied Physics Laboratory and contract NASW-00002 to the Carnegie Institution of Washington. R.J.V., R.M.K., and A.L.S. are supported by the MESSENGER Participating Scientist Program.

22 April 2008; accepted 28 May 2008
10.1126/science.1159467

Reduced Responses to Selection After Species Range Expansion

Benoit Pujol and John R. Pannell*

Range expansion reduces genetic variation in populations at the margins of species' geographic distributions as a result of colonization bottlenecks (*1*, *2*). As a result, range expansion should compromise the adaptive potential of marginal populations (*3*). We tested this prediction by using selection experiments in multiple populations of the European plant *Mercurialis annua*, which expanded its range into Spain and Portugal along the eastern and western coasts of the Iberian Peninsula after the Pleistocene glaciation (*4*).

Populations of *M. annua* on both the eastern and western coasts of the Iberian Peninsula showed molecular genetic diversity that was substantially lower than those in the species' putative Pleistocene refugium in North Africa (Tukey-Kramer tests $P < 0.001$ for both comparisons) (Fig. 1A) (*4*). We predicted that heritable variation in phenotypic traits would mirror this pattern and that Iberian populations would consequentially be less responsive to selection than populations in North Africa. To test this prediction, we compared responses to selection on the amount of pollen produced by hermaphrodites in each of five populations of *M. annua* in both eastern and western Iberia with that in each of six populations in North Africa. Pollen production is an important trait that strongly influences gene flow within and between populations (*5*, *6*).

After an initial generation of acclimatization, we selected for both increased and decreased pollen production by hermaphrodites (*7*). We first measured the male reproductive effort (MRE, the biomass of male flowers and their pollen divided by the total above-ground plant biomass) of 60 individual hermaphrodites for each population and found that the six populations sampled from the putative refugium were significantly more variable in their MREs than the five populations from eastern and western corridors (Tukey-Kramer tests $P = 0.029$ and 0.033, respectively) (Fig. 1B). Patterns of variation in plant biomass were similar to those for MRE (fig. S3A), but we found no regional differences in the variability of female reproductive effort (FRE) (fig. S3B).

For each population, we ranked hermaphrodites in terms of the size of their male flowers and inflorescences, we segregated the upper third and the lower third of ranked individuals into separate subpopulations, and we allowed individuals to mate within their respective subpopulations. Although the upper and the lower lines differed in terms of the MRE, as selected (analysis of variance $P < 0.001$) they did not differ in their biomass ($P = 0.39$) or in their FRE ($P = 0.11$). Thus, there was little opportunity for direct selection on either of these traits.

In the progeny of the selected individuals, we found that divergence in MRE between the high and low lines was greater in North African than in either eastern or western Iberian populations, as predicted (Tukey-Kramer tests $P < 0.001$ for both comparisons) (fig. S4). Plant biomass (which is a denominator in the calculation of MRE) showed a correlated pattern of divergence in the reverse direction to that of MRE (Pearson correlation across all 16 populations $r = -0.61$, $P = 0.012$) (fig. S4), but there was no correlated effect on FRE ($r = 0.09$, $P = 0.74$). Additive genetic variance in MRE, calculated by application of the breeder's equation (*7*), was significantly higher in North Africa than in both eastern and western Iberia (Tukey-Kramer tests $P < 0.001$ for both comparisons) (Fig. 1C).

Our study provides evidence that range expansion can compromise the ability of a species to respond to selection in marginal populations by depleting their additive genetic variation. Interestingly, a study of mosquito populations (*8*) in southeastern North America found that additive genetic variation in developmental traits actually increased toward the species' range margin. This contrasting pattern may be explained by short-term conversion of nonadditive to additive components of genetic variation after bottlenecks (*9*). However, if such a process occurred at all in *M. annua*, the effect was only transient. Certainly, the depletion of heritable variation in marginal populations is likely to be a common feature in species that have expanded their range by passing through genetic bottlenecks (*3*). Such populations may face a higher risk of extinction if they are less able to respond to selection under a changing environment.

References and Notes

1. R. J. Petit *et al.*, *For. Ecol. Manage.* **156**, 49 (2002).
2. C. G. Eckert, K. E. Samis, S. C. Lougheed, *Mol. Ecol.* **17**, 1170 (2008).
3. G. Hewitt, *Nature* **405**, 907 (2000).
4. D. J. Obbard, S. A. Harris, J. R. Pannell, *Am. Nat.* **167**, 354 (2006).
5. T. R. Meagher, *Am. Nat.* **137**, 738 (1991).
6. R. A. Ennos, *Heredity* **72**, 250 (1994).
7. Materials and methods are available on *Science* Online.
8. P. Armbruster, W. E. Bradshaw, C. M. Holzapfel, *Evol. Int. J. Org. Evol.* **52**, 1697 (1998).
9. Y. Naciri-Graven, J. Goudet, *Evol. Int. J. Org. Evol.* **57**, 706 (2003).

10. We thank S. Barrett, M. Dorken, N. Harberd, B. Hill, C. Hughes, and J. Langdale for comments on the manuscript. The work was funded by a grant to J.R.P. from the Natural Environment Research Council, UK.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/96/DC1
Materials and Methods
SOM Text
Figs. S1 to S4
References

10 March 2008; accepted 12 May 2008
10.1126/science.1157570

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

*To whom correspondence should be addressed. E-mail: john.pannell@plants.ox.ac.uk

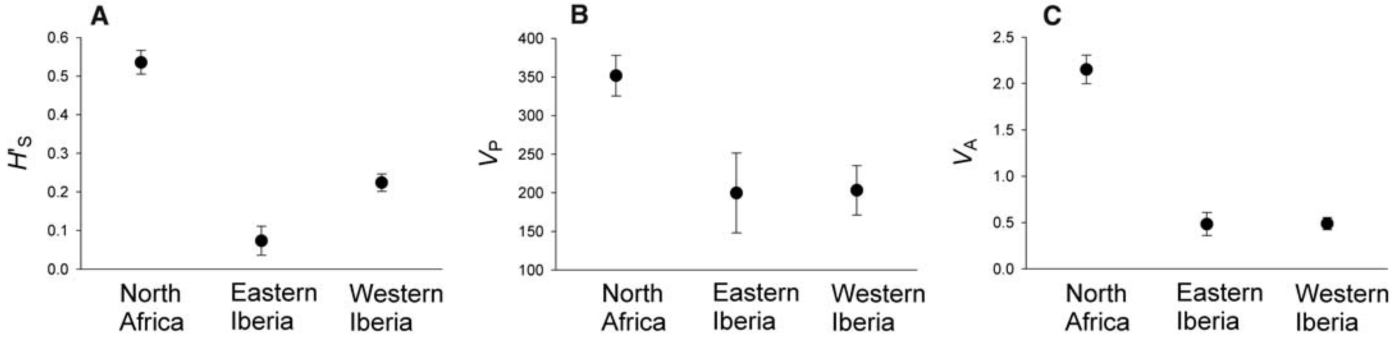


Fig. 1. Estimates of (**A**) neutral genetic diversity, (**B**) phenotypic variance in MRE, and (**C**) additive genetic variance in MRE of *M. annua* for North African, eastern Iberian, and western Iberian populations. Neutral diversity within populations, H'_S, was estimated by Obbard *et al*. (*4*). Error bars are means ± SEM.

Phanerozoic Trends in the Global Diversity of Marine Invertebrates

John Alroy,[1]* Martin Aberhan,[2] David J. Bottjer,[3] Michael Foote,[4] Franz T. Fürsich,[5] Peter J. Harries,[6] Austin J. W. Hendy,[7,8] Steven M. Holland,[9] Linda C. Ivany,[10] Wolfgang Kiessling,[2] Matthew A. Kosnik,[11] Charles R. Marshall,[12] Alistair J. McGowan,[13] Arnold I. Miller,[7] Thomas D. Olszewski,[14] Mark E. Patzkowsky,[15] Shanan E. Peters,[4,16] Loïc Villier,[17] Peter J. Wagner,[11] Nicole Bonuso,[3,18] Philip S. Borkow,[19] Benjamin Brenneis,[2] Matthew E. Clapham,[3,20] Leigh M. Fall,[14] Chad A. Ferguson,[7] Victoria L. Hanson,[4,9] Andrew Z. Krug,[4,15] Karen M. Layou,[7,9,21] Erin H. Leckey,[22] Sabine Nürnberg,[2] Catherine M. Powers,[3] Jocelyn A. Sessa,[7,15] Carl Simpson,[4,23] Adam Tomašových,[4,24] Christy C. Visaggi[10,25]

It has previously been thought that there was a steep Cretaceous and Cenozoic radiation of marine invertebrates. This pattern can be replicated with a new data set of fossil occurrences representing 3.5 million specimens, but only when older analytical protocols are used. Moreover, analyses that employ sampling standardization and more robust counting methods show a modest rise in diversity with no clear trend after the mid-Cretaceous. Globally, locally, and at both high and low latitudes, diversity was less than twice as high in the Neogene as in the mid-Paleozoic. The ratio of global to local richness has changed little, and a latitudinal diversity gradient was present in the early Paleozoic.

Diversity curves showing changes through the geological record in the number of fossil marine genera or families have fueled decades of macroevolutionary research (*1*–*10*). Traditionally, these curves were based on literature compilations that recorded only the first and last appearances of taxa (*3*, *5*). These compilations suggested that diversity rapidly rose during the Cambrian and Ordovician and then either stayed at a plateau (*3*) or declined erratically (*5*) through the Paleozoic. There was a modest rebound after the end-Permian mass extinction, the largest (*11*) and most ecologically important (*9*, *12*) of the Phanerozoic. Diversity in these curves then rose steadily and with a possibly increasing absolute rate, suggesting to some an exponential radiation (*4*, *6*, *9*, *10*).

The appearance of a Paleozoic equilibrium followed by a nearly unbridled Meso-Cenozoic radiation has presented a puzzle: How could global diversity reach limits and then much later cast them off, rising to far higher levels than those seen during the Paleozoic? The possibility that the apparent radiation was exaggerated by secular trends in the quality and especially quantity of the preserved fossil record (*8*, *13*–*15*) was first proposed more than three decades ago (*16*). This claim was at first put aside because sampling effects were thought to be minor and could not be assessed without more detailed information (*2*, *8*).

The Paleobiology Database (http://paleodb.org) makes it possible to address the problem with the use of contemporary statistical methods because it records occurrences of genera and species within particular fossil collections. An intensive data collection effort has quadrupled the database since data for two long Phanerozoic intervals were presented in 2001 (*17*). This initiative has focused on both filling gaps and sampling the Cenozoic at a high level (*18*). The data set includes 44,446 collections with individually recorded ages and geographic coordinates. The collections comprise 284,816 fossil occurrences of 18,702 genera that equate to ~3.5 million specimens and derive from 5384 literature sources.

Sampling and counting. The amount of data per time interval and, therefore, the shape of a diversity curve may vary greatly as a result of uneven preservation and sampling effort. The key advantage of collection data is that this variation can be removed by subsampling (*17*). A random subset of the available collections is drawn until each interval, called a sampling bin, includes the same estimated number of specimens. Genera are counted, and the procedure is repeated to obtain averages. We tallied actual specimen counts when available and otherwise estimated them using a gently curved, one-parameter empirical function that relates the logarithms of specimen and genus counts in each collection (*18*). The parameter is called a calibrated weight. Previous studies (*17*, *19*–*22*) all presumed that this relation was log-log linear and had the same shape in every time interval, or else that there was little change through time in the average size of collections. We instead rarefied actual abundances to produce a separate estimation curve for each interval.

Additionally, we weighted the chance of drawing each collection inversely by its specimen count to distribute sampling both spatially and environmentally, which avoids underestimation of global diversity (*18*). We excluded samples from entirely unlithified rocks, sieved samples from poorly lithified rocks, and samples that preserve original aragonite because it is easier to collect small and fragile specimens in such cases. Furthermore, samples falling into any of these categories are extremely uncommon before the Cenozoic, and the Cenozoic samples are concentrated in a narrow region of the temperate zone [supporting online material (SOM) text]. Finally, before each round of subsampling we restricted the data set to 65 randomly drawn references per interval, exceeding this figure only when more are needed to provide the quota of specimens used in subsampling. Use of a reference quota holds the effective size of the sampling universe more constant, which avoids such problems as a correlation between apparent diversity and the geographic extent of fossil collections (SOM text).

A second issue is how to tally genera (*17*). Conventional protocols count not only genera sampled inside a temporal bin but also genera found at any time before and after a bin (but not inside it). Adding these unsampled taxa to the count creates dropoffs at the edges of curves,

[1] National Center for Ecological Analysis and Synthesis, University of California–Santa Barbara, 735 State Street, Santa Barbara, CA 93101, USA. [2] Museum für Naturkunde der Humboldt-Universität zu Berlin, Invalidenstrasse 43, D-10115 Berlin, Germany. [3] Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089–0740, USA. [4] Department of the Geophysical Sciences, University of Chicago, Chicago, IL 60637, USA. [5] Geozentrum Nordbayern, Friedrich-Alexander Universität Erlangen, Loewenichstrasse 28, D-91054, Germany. [6] Department of Geology, University of South Florida, Tampa, FL 33620–5201, USA. [7] Department of Geology, University of Cincinnati, Cincinnati, OH 45221–0013, USA. [8] Department of Geology and Geophysics, Yale University, New Haven, CT 06520–8109, USA. [9] Department of Geology, University of Georgia, Athens, GA 30602–2501, USA. [10] Department of Earth Sciences, Syracuse University, Syracuse, NY 13244, USA. [11] Department of Paleobiology, Smithsonian Institution, Washington, DC 20013–7012, USA. [12] Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA. [13] Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK. [14] Department of Geology and Geophysics and Faculty of Ecology and Evolutionary Biology, Texas A&M University, College Station, TX 77843–3115, USA. [15] Department of Geosciences, Pennsylvania State University, University Park, PA 16802–2714, USA. [16] Department of Geology and Geophysics, University of Wisconsin–Madison, Madison, WI 53706, USA. [17] Centre de Sédimentologie-Paléontologie, Université de Provence, 13 331 Marseille Cedex 3, France. [18] Department of Geological Sciences, California State University, Fullerton, CA 92834–6850. [19] Science Department, Hamilton Local School District, Columbus, OH 43207, USA. [20] Department of Earth and Planetary Sciences, University of California–Santa Cruz, Santa Cruz, CA 95064, USA. [21] Department of Geology, College of William and Mary, Post Office Box 8795, Williamsburg, VA 23187, USA. [22] Department of Geological Sciences, University of Colorado, Boulder, CO 80309–0399, USA. [23] Department of Biology, Duke University, Durham, NC 27708–0338, USA. [24] Geological Institute, Slovak Academy of Sciences, Dúbravská cesta 9, 84005, Bratislava 45, Slovakia. [25] Department of Biology and Marine Biology, University of North Carolina–Wilmington, Wilmington, NC 28403.

*To whom correspondence should be addressed. E-mail: alroy@nceas.ucsb.edu

local depression of curves near extinction and origination events [the Signor-Lipps effect (*23*)], and interpolated presences of polyphyletic ("wastebasket") genera that have artificially long ranges with large gaps connecting unrelated species. Our direct documentation of occurrences allowed us to avoid all such problems by counting only genera that have actually been sampled. We used a variant of this sampled-in-bin method (*20*) that corrects for residual error by assessing the proportion of genera found immediately before and after a sampling bin but not inside it. This correction has little effect other than reducing some short-term variation (*18*).

Global diversity. The sampling-standardized diversity curve (Fig. 1) shows many key features of older curves (e.g., *5*), such as the Cambro-Ordovician radiation, 78% end-Permian extinction, and 63% end-Triassic extinction (*18*). However, it also includes features that are not highly visible in earlier published curves. Some of them are short-term excursions that may not be robust, such as the brief peak in our sixth Cretaceous bin. Others, however, are seen in numerous treatments of our data (SOM text). (i) The curve suggests that there was a large mid-Devonian drop with no clear recovery until the Permian, instead of a mid-Paleozoic plateau (*3*, *6*). The initial decline begins well before the Frasnian-Famennian ecological collapse, and diversity does not fall across that boundary. (ii) The onset of the late Paleozoic glacial interval was no earlier than the late Famennian (*24*). However, the curve's large mid-Permian increase roughly corresponds with the end of glaciation and an increase in the number of latitudinally restricted tropical genera (*25*). (iii) The recovery from the Permo-Triassic mass extinction is so rapid that Early Triassic standing diversity is only 32% lower than Late Permian standing diversity. (iv) For similar reasons, the 55% Cretaceous-Tertiary mass extinction does not register as a visible net change in diversity at our curve's level of resolution.

Most importantly, the curve casts doubt on the existence of an exponential radiation extending throughout the Mesozoic and Cenozoic. The Triassic data points are higher than the early Jurassic points instead of lower, and there is little net change from the mid-Cretaceous through the Paleogene. Even the Neogene peak is subdued: The curve's last data point is only a factor of 1.14 and 1.37 higher than the two highest points in the early and mid-Paleozoic (both in the Early Devonian) and a factor of 1.74 higher than the median Paleozoic point. Older subsampling methods suggest an even smaller increase (*18*), and the Neogene values may be exaggerated by geographic factors (SOM text). Thus, the new results suggest that any post-Paleozoic radiation was largely confined to the Jurassic and Early Cretaceous.

In sum, the net increase in global diversity over nearly a half-billion years was proportionately not much larger than some of the changes in genus counts between neighboring 11-million-year (My) intervals. However, some treatments of Sepkoski's genus-level compilation imply that the mid-Paleozoic–to–Neogene increase was by a factor of 3.5 (*7*), or even 4.1 (*6*). We next show that within-collection diversity patterns and changes in latitudinal gradients are only consistent with the new curve.

Collection-level diversity and evenness. Abundance distributions are even when each taxon is represented by a similar number of specimens. Evenness is of intrinsic ecological interest because it controls sampled richness when collections are of the size normally studied by paleoecologists (about 100 to 300 specimens) (*26*–*28*). Fortuitously, the single governing parameter estimated by the calibrated-weights method is easily translated into Hurlbert's probability of interspecific encounter (PIE)

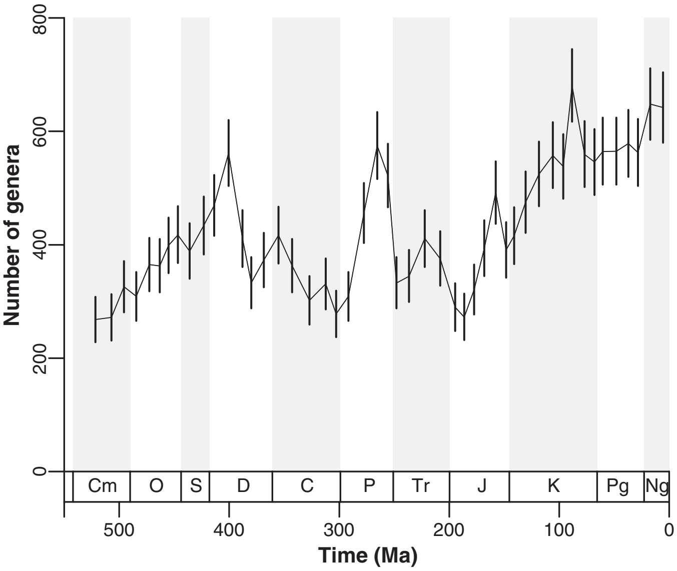


Fig. 1. Genus-level diversity of both extant and extinct marine invertebrates (metazoans less tetrapods) during the Phanerozoic, based on a sampling-standardized analysis of the Paleobiology Database. Points represent 48 temporal bins defined to be of roughly equal length (averaging 11 My) by grouping short geological stages when necessary. Vertical lines show the 95% confidence intervals based on Chernoff bounds, which are always conservative regardless of the number of genera that could be sampled or variation in their sampling probabilities (*18*). Data are standardized by repeatedly drawing collections from a randomly generated set of 65 publications until a quota of 16,200 specimens has been recovered in each bin. On average, 461 collections had to be drawn to reach this total. The curve shows average values found across 20 separate subsampling trials—enough to yield high precision with such large sample sizes. Ma, million years ago. Cm, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; J, Jurassic; K, Cretaceous; Pg, Paleogene; Ng, Neogene.

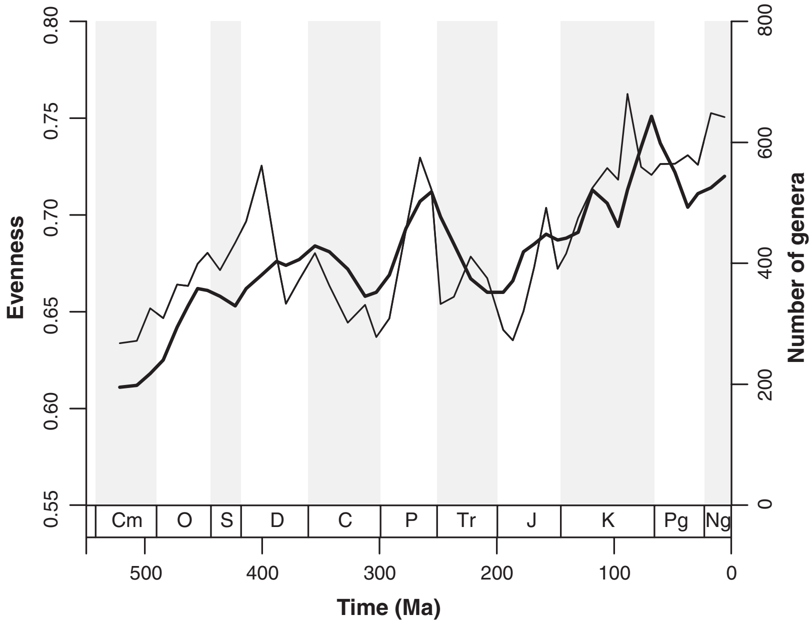


Fig. 2. Evenness estimates (thick line) for 11-My-long bins, based on Hurlbert's PIE index (*27*). The values are computed from the calibrated occurrence weights used in subsampling, and each is a weighted moving averages across five consecutive bins. Changes in global diversity (thin line, same as in Fig. 1) are shown for comparison.

index of evenness (*27*). PIE bears no necessary relation to the total number of taxa that might be sampled.

Evenness changed substantially through the Phanerozoic, if not as dramatically as sometimes suggested (*29*), and the greatest values are seen in the late Cretaceous and Cenozoic (Fig. 2). They imply, for example, that 200 specimens will yield about 11.3 genera in the latest Ordovician, 22.2 in the Paleocene, and 19.3 in the late Neogene. The general pattern of a long-term increase is confirmed by related (but more restricted) studies (*26*, *30*). However, the evenness curve (Fig. 2) does not simply increase. It suggests a plateau between the Late Ordovician and Carboniferous and then a large rise through the Permian. After a weak recovery from the Permo-Triassic decline, it shows little change until a rise in the Late Cretaceous (*18*). All of these features are broadly consistent with the idea that local diversity reached its maximum sometime during the past 100 My (*29*) as the number of occupied niches expanded (*31*). However, they do not suggest a radical increase between the early Paleozoic and Cenozoic. They also do not mirror the shift in abundance distributions from simple to complex shapes at the Permo-Triassic boundary (*12*), because evenness dropped across this boundary instead of rising, and the Triassic figures are well within the narrow range seen throughout most of the Paleozoic.

Theoretically, if biogeographic and environmental gradients (i.e., beta diversity) do not change, global diversity should track local sampled diversity and, therefore, evenness. Furthermore, the global curve is methodologically grounded on the evenness data because the calibrated function estimates fewer specimens per collection when evenness is high, which causes more collections to be drawn. Hence, it is no surprise that after logging and differencing the global and local curves (Fig. 2), we find a significant correlation (Spearman's rank-order correlation ρ = 0.332, P = 0.023). Specific similarities include the parallel increase during the Cambro-Ordovician radiation, the joint rise during the Permian, and the drop in evenness during the severe global extinction at the Permo-Triassic boundary (*9*, *11*, *12*).

However, the curves depart from one another for long stretches of time. Evenness remained relatively high during the long late Paleozoic trough in global richness (Fig. 2). The offset may be driven by geography: The continents were widely dispersed during the early Paleozoic, late Mesozoic, and Cenozoic, and the late Paleozoic low roughly corresponds with the assembly of the supercontinent Pangaea. Thus, steeper biogeographic gradients during the late Phanerozoic may have accommodated greater global diversity (*1*), and the same may have been true during the early Paleozoic. Nevertheless, this hypothesis needs to be explored in more detail because the mid-Phanerozoic offset between the curves is not consistent, and nonbiogeographic factors such as onshore-offshore gradients also may have changed.

Latitudinal diversity gradients. To ensure that the global curve captured any radiation in the tropics that might have occurred, we deliberately focused on the Neogene tropics while simultaneously establishing a baseline sampling level for the entire Phanerozoic. Counts of references in the Paleobiology Database, Sepkoski's compendium, GeoRef, and the Treatise on Invertebrate Paleontology suggest that our data sample the tropics as well as, or better than, those compilations (SOM text). Enough data are available to compare latitudinal belts within several key time intervals.

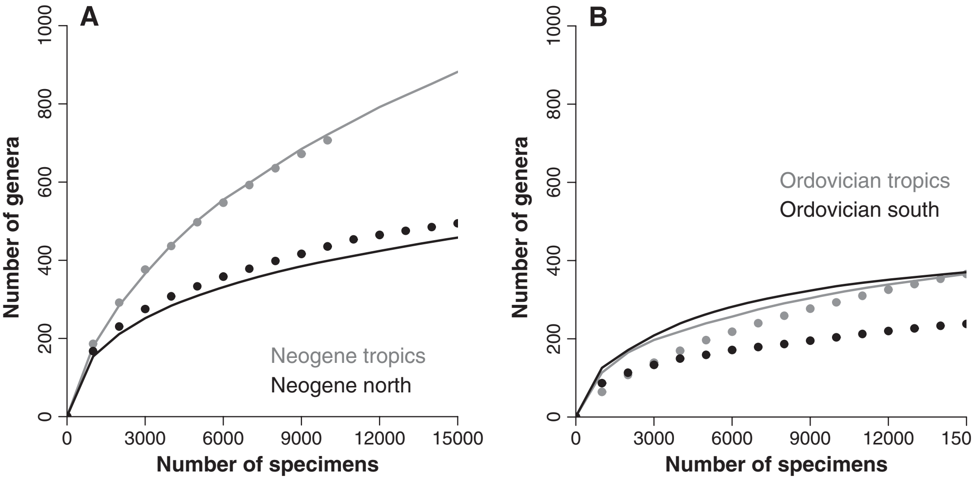
Subsampling curves for individual Neogene bins show that low-latitude diversity (within 30° paleolatitude of the equator) was substantially higher than northern temperate zone diversity (Fig. 3A). The Neogene tropical curves overlap, and the high-latitude curves are also very similar to each other. In the Ordovician, there was little benthic habitat in the north, so we have computed low-latitude and southern temperate zone curves (Fig. 3B). They show a clear latitudinal diversity gradient for one bin but not the other. Thus, regardless of whether the modern gradient came into existence recently, at least some much older intervals did witness similar patterns. Furthermore, because there is a large difference between the Neogene and Ordovician at both high and low latitudes (Fig. 3), the moderate net increase through the Phanerozoic (Fig. 1) seems to have been a global phenomenon instead of being driven strictly by a radiation in the tropics.

Previous data sets. The fact that most treatments of older compilations depict a massive Cretaceous-Cenozoic radiation (*1*–*7*, *9*, *10*) raises the question of whether the differences are primarily methodological or primarily related to coverage of the literature. We answer this question by tabulating our data set and Sepkoski's range-based, genus-level compendium (*5*) in exactly the same way. Because it is impossible to either sample-standardize or take sampled-in-bin counts if only ranges are available (*5*), we use raw, unstandardized data and treat our genus age ranges as if they were continuously sampled.

The two data sets yield similarly shaped curves of a comparable magnitude (Fig. 4). The genus-level curves (Fig. 4) virtually overlap, and the same is true for ordinal-level data sets (SOM text). The new genus-level curve is higher than Sepkoski's throughout much of the Mesozoic but still suggests a large Cenozoic radiation, albeit smaller than that in Sepkoski's data. Our raw curve suggests a 3.74 times difference between the late Neogene and the median Ordovician, Silurian, or Devonian interval—slightly more than the factor of 3.54 seen in Sepkoski's compendium (Fig. 4). Both curves identify not just major but also minor features, such as a peak in the late Jurassic.

Thus, the dramatic differences between the standardized and conventional curves (Figs. 1 and 4) do not result from data-quality or -quantity

Fig. 3. Low- (30°S to 30°N paleolatitude) and high-latitude subsampling curves for individual 11-My-long bins. Gray lines indicate low-latitude data. (**A**) Data for the Cenozoic bins, including the Early/Middle Miocene (dotted lines) and Late Miocene/Pliocene/Pleistocene (solid lines). Black lines indicate data from above 30°N. (**B**) Data for the Ordovician bins, including the Llanvirn (dotted lines) and Caradoc (solid lines). Black lines denote data from below 30°S.



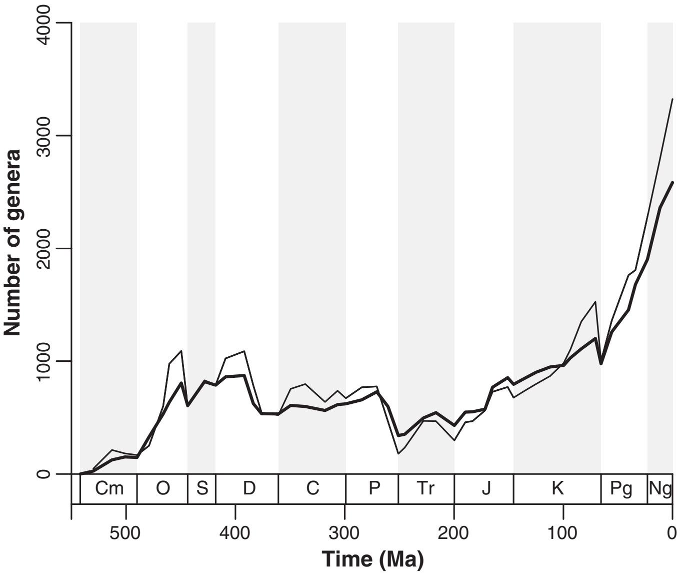


Fig. 4. Genus-level diversity curves based on Sepkoski's compendium [thin line (*5*)] and our new data (thick line). Counts are of marine metazoan genera crossing boundaries between temporal bins (boundary crossers) and exclude tetrapods. Ranges are pulled forward from first fossil appearances to the Recent, instead of ending at the last known fossil appearance. Extant genera are systematically marked as such based on Sepkoski's compendium and the primary literature. There is no correction for sampling, and genera are assumed to be sampled everywhere within their ranges because Sepkoski's traditional synoptic data (*5*) do not record occurrences within individual collections.

problems. We instead attribute these differences to two biases that can only be removed with standardized sampling and sampled-in-bin counting of occurrence data (SOM text). First, we have exceeded the magnitude of Sepkoski's curve in the mid-Phanerozoic with a modest data set but fallen just short of it in the Cenozoic with a massive data set (*18*). Fully matching his curve would require us to make our sampling even more heterogeneous. Hence, the literature compiled by Sepkoski seems likely to contain a strong Cenozoic sampling bias. Second, counting each extant taxon as sampled everywhere from its last fossil appearance to the Recent [i.e., the Pull of the Recent (*7, 32*)] exaggerates the artifactual Cretaceous and Cenozoic increase. This problem cannot affect counts of sampled taxa because they take no note of which genera are extant or which extend beyond a particular bin.

Conclusion. The new diversity curve (Fig. 1) records substantial volatility, including some potentially meaningful excursions that might relate to evolutionary innovations, paleogeographic shifts, global climate change, sea-level change, or other factors. In particular, the fact that local and global diversity have not always changed in tandem (Fig. 2) implies that compositional differences between environments or geographic regions have waxed and waned.

Regardless of whether this is true, a more general and important pattern is evident: Most of the Meso-Cenozoic radiation took place well before the Late Cretaceous and Cenozoic, and the net increase through the Phanerozoic was proportionately small relative to the enormous amount of time that elapsed. Strong latitudinal diversity gradients as far back as the Ordovician and modest changes in local-scale sampled diversity (Fig. 2) (*26*, *29*, *30*) are consistent with the suggestion that global biodiversity is constrained. Additional evidence includes rapid rebounds from all of the major extinction episodes (Fig. 1) (*3*).

Although any limit to diversity could not have been static and must have had a net increase (*3*, *13*, *16*, *17*, *33*), our results cannot be reconciled with previous studies that argued for exponential long-term growth on the basis of raw, unstandardized data (*1*, *4*, *6*, *9*, *10*). Allowing for ecological reorganizations in the wake of mass extinction (*9*) or for the addition of ecological niches by means of evolutionary innovation (*31*, *34*) does not explain how diversification could have been limitless in the face of interval-to-interval changes that rivaled the entire net increase over the Phanerozoic.

Thus, we now must ask what mechanisms could have led to saturation. They may have involved the way that energy is captured from lower trophic levels. This capture could have remained roughly constant (*35*). Alternatively, if total energy capture did increase (*36*), this change may have been offset by the diversification of groups with high metabolic rates (*9*), making it energetically difficult for a large net radiation to occur.

References and Notes

1. J. W. Valentine, *Palaeontology* **12**, 684 (1969).
2. J. J. Sepkoski Jr., R. K. Bambach, D. M. Raup, J. W. Valentine, *Nature* **293**, 435 (1981).
3. J. J. Sepkoski Jr., *Paleobiology* **10**, 246 (1984).
4. M. J. Benton, *Science* **268**, 52 (1995).
5. J. J. Sepkoski Jr., in *Global Events and Event Stratigraphy*, O. H. Walliser, Ed. (Springer, Berlin, 1996), pp. 35–52.
6. R. K. Bambach, *Geobios* **32**, 131 (1999).
7. M. Foote, *Paleobiology* **26**, (suppl. 1), 74 (2000).
8. A. I. Miller, *Paleobiology* **26**, (suppl. 1), 53 (2000).
9. R. K. Bambach, A. H. Knoll, J. J. Sepkoski Jr., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6854 (2002).
10. S. M. Stanley, *Paleobiology* **33**, (suppl. to no. 4), 1 (2007).
11. D. M. Raup, J. J. Sepkoski Jr., *Science* **215**, 1501 (1982).
12. P. J. Wagner, M. A. Kosnik, S. Lidgard, *Science* **314**, 1289 (2006).
13. D. M. Raup, *Paleobiology* **2**, 289 (1976).
14. S. E. Peters, M. Foote, *Paleobiology* **27**, 583 (2001).
15. J. S. Crampton *et al*., *Science* **301**, 358 (2003); published online 12 June 2003, 10.1126/science.1085075.
16. D. M. Raup, *Science* **177**, 1065 (1972).
17. J. Alroy *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6261 (2001).
18. Materials and methods are available as supporting material on *Science* Online.
19. J. Alroy, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **127**, 285 (1996).
20. A. I. Miller, M. Foote, *Paleobiology* **22**, 304 (1996).
21. J. Alroy, *Paleobiology* **26**, 707 (2000).
22. A. M. Bush, M. J. Markey, C. R. Marshall, *Paleobiology* **30**, 666 (2004).
23. P. W. Signor, J. H. Lipps, *Geol. Soc. Am. Spec. Pap.* **190**, 291 (1982).
24. M. Streel, M. V. Caputo, S. Loboziak, J. H. G. Melo, *Earth Sci. Rev.* **52**, 121 (2000).
25. M. G. Powell, *Geology* **33**, 381 (2005).
26. M. G. Powell, M. Kowalewski, *Geology* **30**, 331 (2002).
27. T. D. Olszewski, *Oikos* **104**, 377 (2004).
28. S. E. Peters, *Paleobiology* **30**, 325 (2004).
29. A. M. Bush, R. K. Bambach, *J. Geol.* **112**, 625 (2004).
30. M. Kowalewski *et al*., *Paleobiology* **32**, 533 (2006).
31. R. K. Bambach, in *Biotic Interactions in Recent and Fossil Benthic Communities*, M. J. S. Tevesz, P. L. McCall, Eds. (Plenum, New York, 1983), pp. 719–746.
32. D. M. Raup, *Bull. Carnegie Mus. Nat. Hist.* **13**, 85 (1979).
33. G. G. Simpson, in *Evolution After Darwin*, *Vol. I*, *The Evolution of Life*, S. Tax, Ed. (Univ. of Chicago Press, Chicago, 1960), pp. 117–180.
34. W. I. Ausich, D. J. Bottjer, *Science* **216**, 173 (1982).
35. L. Van Valen, *Evol. Theory* **1**, 179 (1973).
36. R. K. Bambach, *Paleobiology* **19**, 372 (1993).
37. We thank R. Bambach and D. Jablonski for extensive and helpful discussions and M. Apel, M. Bosnakoff, D. Buick, K. Bulinski, M. Gibson, N. Heim, M. Hethke, D. Korn, K. Koverman, J. Millet, K. Nell, P. Novack-Gottshall, T. Schossleitner, C. Schweitzer, M. Uhen, P. Wall and especially D. Hempfling, S. Hicks, and U. Merkel for contributing to the data set. This work was funded by the NSF Division of Environmental Biology (grant DEB-0083893 to C.R.M., J.A., A.I.M., J. M. Hayes, and D. H. Rothman), the NSF Sedimentary Geology and Paleontology program (grants EAR-0521425 and EAR-0545059 to J.A., EAR-0434443 to D.J.B., and EAR-0105609 to M.F.), the NASA Program in Exobiology and Evolutionary Biology (grants to A.I.M. and M.E.P.), and the VolkswagenStiftung (grant to W.K.). Our research was conducted while J.A. was a Center Associate at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (grant DEB-0553768), the University of California–Santa Barbara, and the state of California. This is Paleobiology Database publication number 80.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/97/DC1
Materials and Methods
SOM Text
Figs. S1 to S21
References

25 February 2008; accepted 2 June 2008
10.1126/science.1156963

Self-Sorted, Aligned Nanotube Networks for Thin-Film Transistors

Melburne C. LeMieux,[1] Mark Roberts,[1] Soumendra Barman,[1] Yong Wan Jin,[2] Jong Min Kim,[2] Zhenan Bao[1]*

To find use in electronics, single-walled carbon nanotubes need to be efficiently separated by electronic type and aligned to ensure optimal and reproducible electronic properties. We report the fabrication of single-walled carbon nanotube (SWNT) network field-effect transistors, deposited from solution, possessing controllable topology and an on/off ratio as high as 900,000. The spin-assisted alignment and density of the SWNTs are tuned by different surfaces that effectively vary the degree of interaction with surface functionalities in the device channel. This leads to a self-sorted SWNT network in which nanotube chirality separation and simultaneous control of density and alignment occur in one step during device fabrication. Micro-Raman experiments corroborate device results as a function of surface chemistry, indicating enrichment of the specific SWNT electronic type absorbed onto the modified dielectric.

The combination of superior electrical and mechanical properties in single-walled carbon nanotubes (SWNTs) (*1*, *2*) continues to advance applications, including flexible electronics (*3*, *4*), biosensors and biochemical sensors (*5*), and solar cell technology (*6*). Despite enormous progress toward the potential applied uses of SWNT devices, such applications will not be realized unless fundamental issues concerning the controlled reproducible placement, alignment, and separation based on chirality and/or diameter can be solved.

Accurate orientation of SWNTs has been achieved from high-temperature growth (*7*), and on a limited scale, at ambient conditions with diverse approaches including dielectrophoresis (*8*), gas-flow (*9*), evaporating droplet (*10*, *11*), and absorption affinity patterns (*12*, *13*). As a more realistic route for integration, researchers have recently turned to random nanotube networks that are easily absorbed or deposited from solution to form two-dimensional (2D) nanotube films (*4*, *9*, *14*, *15*) as the active semiconducting layer in SWNT network thin-film transistors (Fig. 1A). However, these devices suffer from poor on/off ratios due to the presence of both semiconducting and metallic tubes intrinsic to all SWNT synthesis methods. Although metallic SWNTs may serve as high-performance interconnects (*16*), they lead to high off current and shorted transistors. The metallic SWNTs may be burned off after fabrication (*17*), but this additional processing step is neither well controlled nor scalable.

Our goal is to move beyond random nanotube films to submonolayer SWNT networks with controlled topology in which chirality, alignment, and density may be tuned or sorted, ideally during fabrication. We took SWNT solutions and spincoated them onto piranha-cleaned silicon wafers with a 300-nm dry thermal oxide (*18*) modified with various silane monolayers chosen for their relative ease of surface modification and the variety of chemical functional groups available (Fig. 1). Independent of the solvent volatility used for deposition, if the SWNT solution is spincoated on bare (unmodified) SiO_2, no nanotube absorption occurs. To control absorption characteristics, we functionalized the SiO_2 dielectric surface with amine- and phenyl-terminated silanes (Fig. 1). The thickness of the amine and phenyl surfaces, measured by ellipsometry, was 0.7 and 0.4 nm, respectively, and root mean square (RMS) surface roughness (*19*) did not exceed 0.2 nm (table S1), indicating uniform monolayers.

Aminosilanes are used because they enhance the absorption of nanotubes (*20*) and improve the performance of isolated SWNT thin-film transistors (TFTs) as sensors (*21*). Furthermore, the pioneering work by Papadimitrakopoulos has shown selective absorption of amines toward semiconducting SWNTs (*22*, *23*). However, theory (*24*) and experiment (*25*) have shown that aromatic molecules like the phenyl-terminated silane used here interact and bind selectively to metallic SWNTs. The selectivity is attributed to the fact that SWNTs are extended π-electron systems that can interact with other π-electron systems via π-π stacking. This is expected to be stronger with the metallic nanotubes because of a

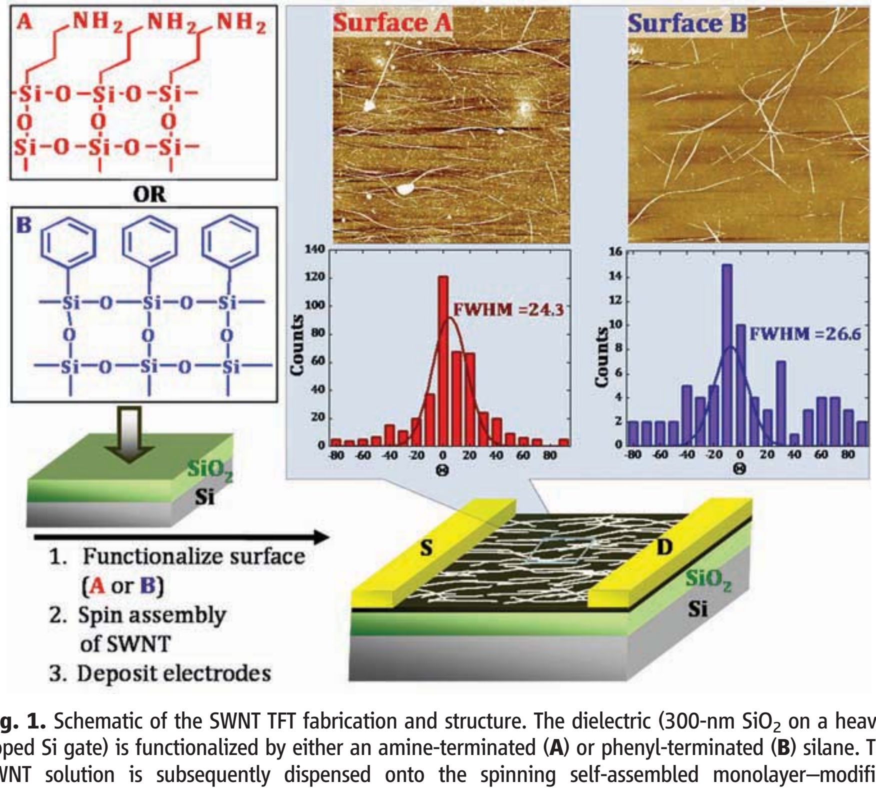


Fig. 1. Schematic of the SWNT TFT fabrication and structure. The dielectric (300-nm SiO_2 on a heavily doped Si gate) is functionalized by either an amine-terminated (**A**) or phenyl-terminated (**B**) silane. The SWNT solution is subsequently dispensed onto the spinning self-assembled monolayer–modified substrate and dried, followed by source (S) and drain (D) gold electrode deposition. Shown here is the top-contact device structure, although similar results are obtained in bottom-contact layouts where source and drain electrodes are deposited before spin assembly of the nanotubes. Upon spincoating, AFM tapping-mode topography images (10 μm by 10 μm, *z* scale = 10 nm) of the nanotubes applied under identical conditions on amine (top) and phenyl (bottom) surfaces reveals that density and alignment, represented by histogram (Θ is angle, in degrees, of variation from an arbitrary direction) below the corresponding AFM images, are a direct function of surface chemistry.

[1] Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. [2] Samsung Advanced Institute of Technology, Giheung-gu, Younggin-si Gyunggi-do 449-712, South Korea.

*To whom correspondence should be addressed. E-mail: zbao@stanford.edu

larger polarizability relative to the semiconducting nanotubes (*8*). An unresolved question, especially relevant for device applications, is whether this selectivity can be translated to interactions at surfaces.

Qualitatively, spincoating the SWNT solution on the various silane-treated surfaces under the same conditions resulted in the tube density being substantially modulated by surface chemistry. As determined from atomic force microscopy (AFM) images (10 μm by 10 μm) (Fig. 1) taken at random locations on the wafer (except at the center of the wafer), the absorption density of the isolated SWNT was higher on aminosilanes (8 to 10 per μm^2) than on the phenyl surfaces (1 to 2 per μm^2). This is expected on the basis of simulation results (*24*) that found that the highest equilibrium absorption energy of an NH_2 group with SWNTs, regardless of chirality, is 0.57 eV, almost three times as strong as that of phenyl groups, which is ~0.20 eV. Upon spincoating, the average alignment over the entire wafer (2.5 cm), as measured by AFM images (10 μm by 10 μm) (Fig. 1), indicates that at any location, ~71% of the SWNTs are aligned within ±10% of an arbitrary axis, with a slightly higher degree of alignment observed on the amine surface as compared to the phenyl surface (histograms in Fig. 1). The hydrodynamic flow, which has been shown to efficiently align nanoparticles on a drying surface (*26*), appears to be markedly enhanced by the spinning wafer and effectively aligns the nanotubes as the solvent evaporates and thins (*10*). The degree of alignment can be tuned on the basis of spin-assembly conditions (fig. S1). The alignment is radial from the center, with the only unaligned area being the portion concentric with the vacuum chuck on the spin coater (~2 mm in diameter on a 12.7-mm-diameter wafer) (fig. S2). No alignment was observed when the SWNTs were deposited by wafer soaking or dropping the solution followed by spinning (fig. S3). Spincoating conditions (volume, rate) could be controlled so that nearly the same SWNT density was absorbed onto amine and phenyl surfaces for comparing the resulting electronic and micro-Raman spectroscopy analysis.

To determine whether and how different surface functionalities affect the electronic response of the SWNT thin-film transistors, we performed electronic testing and micro-Raman mapping in the channel region of the TFTs over random substrate locations. The SWNT TFTs on aminosilane surfaces exhibit p-type behavior with on/off ratios as high as 900,000 and without any postprocessing or device level burn-off (*17*), indicating that the nanotube network is composed primarily of semiconducting nanotubes (Fig. 2, A and B). The average on/off ratio for working devices was ~200,000, and leakage current was in the picoamp range with no dependence on applied gate voltage. Field-effect mobility (*27*) of the TFTs on amine surfaces ranged from ~0.5 to 6 cm^2/V·s (fig. S4). Overall, 38% of the 80 devices tested on amine surfaces had on/off ratios of >100,000, 69% had ratios of >10,000, and more than 95% had ratios of >100 on bottom and top contact devices possessing a channel length of ≥15 μm (*18*). Even at a smaller channel length of 10 μm, an average on/off ratio approaching 10,000 was maintained (fig. S5). Although amines can slightly modulate SWNT doping levels, the on/off ratio is not affected (*21*). Similarly, we confirmed that the high on/off ratios were not due to dedoping by amine groups, because the drop-and-spin samples with the same SWNT density on the amine surface showed no selectivity and a low on/off ratio (fig. S3).

SWNT TFTs fabricated on the aromatic surfaces with the same density, however, always led to conducting devices, and all devices had on/off ratios of less than ~2, indicating that the phenyl surfaces are dominated by metallic nanotubes. Figure 2C shows the averaged transfer curve. The resulting error bars are relatively high due to different levels of "off-current" that averaged around −6.0 × 10^{-6}, which are five to six orders of magnitude higher than those of the off-current on the amine surfaces. Furthermore, the overall higher source/drain current (I_{DS}) maximum "on current" for the phenyl surfaces, as opposed to the amine surfaces, may be another indication that the phenyl surfaces are dominated by metallic nanotubes because of the higher current-carrying capacity. This pronounced difference in the electronic nature of the SWNT networks on the two surfaces is observed down to very short

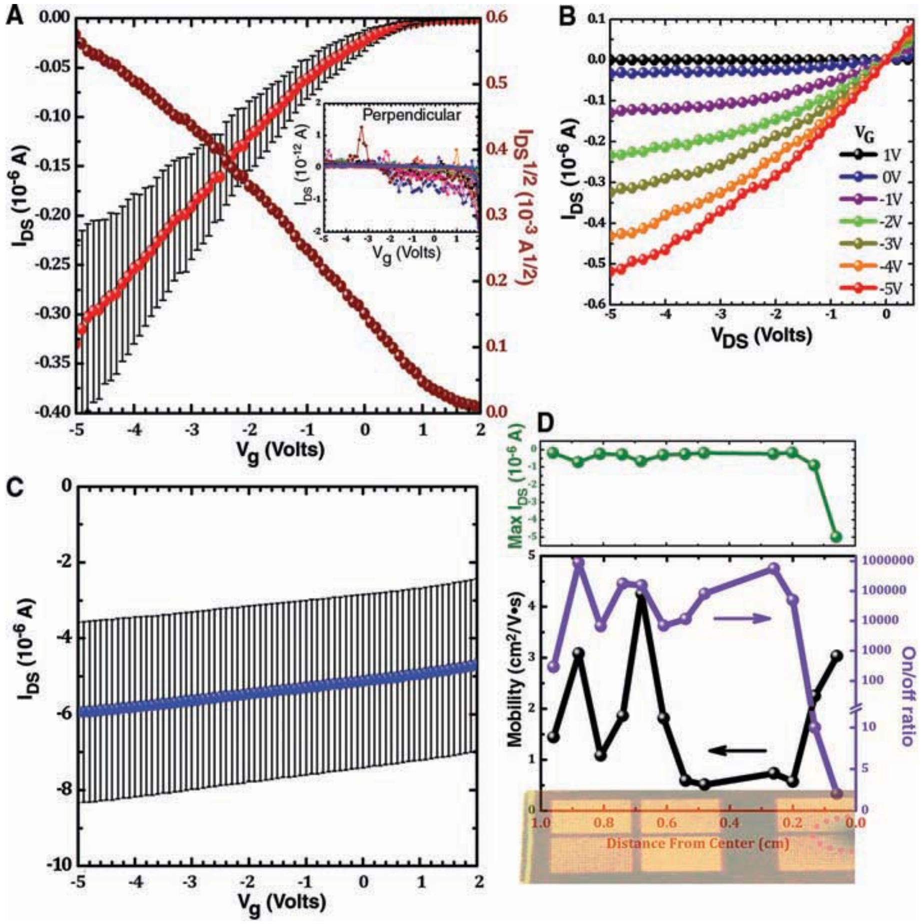


Fig. 2. (**A**) Plot of all measured devices showing the mean ± SD of source/drain current (I_{DS}) as a function of gate voltage (V_g) for SWNT TFTs on an amine-terminated surface showing high on/off ratio in devices with current-flow direction parallel to hydrodynamic flow direction. Because the hydrodynamic flow is radial from the center of the substrate during spincoating, the working devices have current-flow primarily along this direction. Perpendicular to flow, the devices are insulating because a percolating network is suppressed (inset). (**B**) Typical plot of I_{DS} as a function of V_G for the SWNT TFT on amine-terminated surfaces. A slightly positive threshold (V_T ~ 0.7 V) is associated with the transistors as a result of oxygen doping. (**C**) Plot of all measured devices showing the mean ± SD of I_{DS} as a function of V_G for SWNT TFTs on phenyl surfaces showing a high "off" current and a low on/off ratio. Error bars in (A) and (C) represent the SD in the transfer curves. (**D**) Uniformity of the spin-assisted assembly method over a ~2-cm bottom-contact device substrate demonstrating scalability. The plot is superimposed on a device wafer (digital image) with the highlighted shaded circle indicating high (uncontrollable) concentration of SWNTs that is found only in the very center of the wafer (see supporting online material). The top plot represents the maximum "on-current" for one typical wafer, and the bottom plot compares mobility and on/off ratio as a function of test location on the wafer.

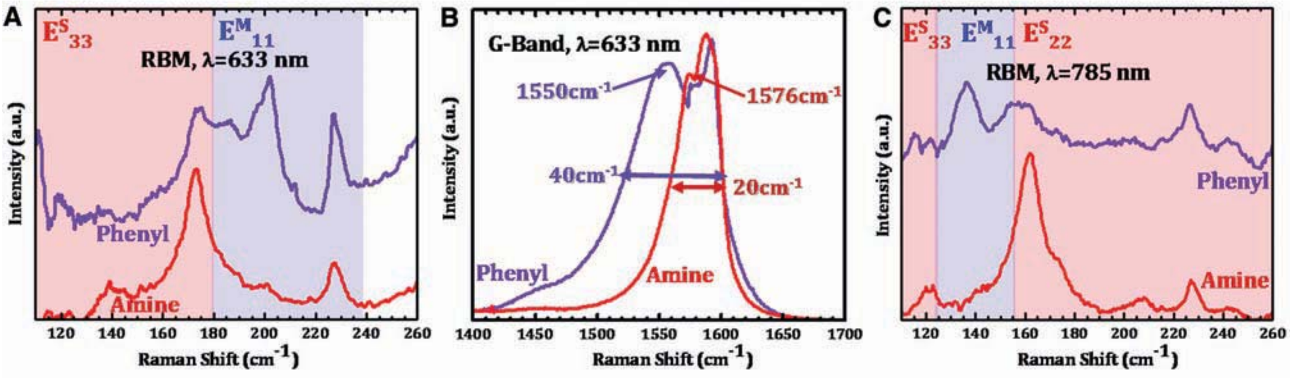


Fig. 3. (**A**) Micro-Raman spectra at 1.96-eV excitation showing RBMs compiled from 12 point maps from at least five different locations of the SWNT network on each surface. All spectra shown here were normalized to the 303-cm^{-1} silicon peak; the peaks below 120 cm^{-1} result from noise, whereas the peak at ~225 cm^{-1} arises from silicon. (**B**) Corresponding G-band compilations at 1.96 eV of each surface shows a marked downshift and broadening of ω_{G^-} band from the SWNTs on the phenyl surface (blue trace) resulting from metallic SWNTs as compared to the amine surface (red trace). The full width at half-maximum of the G$^-$ band is twice as high on the phenyl surface as compared to the amine surface. (**C**) Raman spectra at 1.58-eV excitation showing RBMs compiled from each surface. The phenyl surface (blue trace) is enriched with metallic nanotubes, whereas the amine surface (red trace) is enriched with semiconducting nanotubes.

channel lengths of 2 μm, approaching the length of the SWNTs in this study and indicating a high degree of separation (fig. S6).

To identify tube types (*28*, *29*) absorbed on each surface, we performed micro-Raman mapping over large areas of each sample surface in device channels by using two excitation energies and correlating observed resonant radial breathing modes (RBMs) with chirality using a Kataura plot and a relation between diameter and RBM developed by Dresselhaus *et al*. for isolated SWNTs on SiO_2 (*29*). We used 1.96- and 1.58-eV excitation because most SWNT tubes in this work had a relatively large average diameter but narrow distribution (1.4 ± 0.4 nm as determined by AFM), and should be resonant at these energies. The diameter distribution of arc-discharge nanotubes is known to be quite narrow, more so than that of tubes produced by HiPco and chemical vapor deposition (*30*), allowing this multi-excitation micro-Raman experiment to be highly reliable and thorough. In particular, the 1.96-eV line is resonant with metallic SWNTs (E^M_{11}) and semiconducting SWNTs (E^S_{33}) in a nearly 50/50 ratio (Fig. 3A) and is best suited for separation analysis.

RBM analysis (Fig. 3A and figs. S7, S8, and S10) from compiled spectra shows that tubes absorbed on both surfaces possess the 170-cm^{-1} peak because this corresponds to the average diameter of our arc-discharge SWNTs (fig. S5) (*31*). However, this semiconducting band is clearly much stronger on the amine surface, indicating enrichment of these tubes on amine, as is the 145-cm^{-1} semiconducting peak that is also observed in the reference sample (fig. S7), but completely suppressed in the nanotube network absorbed on the phenyl surface (Fig. 3A).

In contrast, there is strong enrichment of the 200-cm^{-1} peak corresponding to metallic SWNTs on the phenyl surface that is nearly nonexistent on the amine surface. Moreover, the 190-cm^{-1} (metallic band) peak on the phenyl surface, nearly identical in intensity to the 170-cm^{-1} peak, was almost completely eliminated from the amine surface. The corresponding G band (Fig. 3B and figs. S9 and S11) from the nanotube network on each surface differs substantially, indicating that the electronic nature of the absorbed nanotubes is very different. Whereas the amine is characterized by two (G$^+$ and G$^-$) relatively sharp peaks, on the phenyl surfaces, there is a strong downshift of the G$^-$ band accompanied with a distinct Breit-Wigner-Fano line shape characterized by a marked broadening of this same band (Fig. 3B). This downshift and pronounced broadening of the G$^-$ band indicate a strong metallic contribution, and along with the markedly higher integrated area G$^-$/G$^+$ ratio on the phenyl surfaces, are further evidence for a higher ratio of metallic SWNTs on the phenyl surfaces (Fig. 3B). Thus, micro-Raman experiments corroborate the electronic measurements and explain the huge difference in the electrical response of the sorted nanotube network on each surface. RBM analysis from the 1.58-eV line (Fig. 3C) further reinforces this finding because the amine surface is dominated by the 163-cm^{-1} peak (E^S_{22}), whereas the phenyl surface has a strong 134-cm^{-1} peak that corresponds to metallic nanotubes and is highly suppressed on the amine surface.

By spincoating solutions of SWNTs onto functionalized surfaces, we can obtain thin-film transistors with average on/off ratios of >100,000. To consistently achieve such high values from a one-step solution-processed SWNT network TFT without any burnoff or additional processing represents a major step toward the application of nanotube electronics. AFM, electronic, and micro-Raman measurements show that the alignment, density, electronic nature, and chirality of SWNTs can be tuned through careful selection of surface functional groups and the spin-assisted assembly described here.

References and Notes

1. R. H. Baughman, A. A. Zakhidov, W. A. de Heer, *Science* **297**, 787 (2002).
2. J. T. Hu, T. W. Odom, C. M. Lieber, *Acc. Chem. Res.* **32**, 435 (1999).
3. J.-H. Ahn *et al.*, *Science* **314**, 1754 (2006).
4. G. Grüner, *J. Mater. Chem.* **16**, 3533 (2006).
5. E. S. Snow, F. K. Perkins, J. A. Robinson, *Chem. Soc. Rev.* **35**, 790 (2006).
6. M. W. Rowell *et al.*, *Appl. Phys. Lett.* **88**, 233506 (2006).
7. C. Kocabas *et al.*, *Nano Lett.* **7**, 1195 (2007).
8. R. Krupke, F. Hennrich, H. V. Lohneysen, M. M. Kappes, *Science* **301**, 344 (2003).
9. M. D. Lay, J. P. Novak, E. S. Snow, *Nano Lett.* **4**, 603 (2004).
10. R. Duggal, F. Hussain, M. Pasquali, *Adv. Mater.* **18**, 29 (2006).
11. V. V. Tsukruk, H. Ko, S. Peleshanko, *Phys. Rev. Lett.* **92**, 065502 (2004).
12. S. G. Rao, L. Huang, W. Setyawan, S. Hong, *Nature* **425**, 36 (2003).
13. Y. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2026 (2006).
14. E. Artukovic, M. Kaempgen, D. S. Hecht, S. Roth, G. Grüner, *Nano Lett.* **5**, 757 (2005).
15. E. S. Snow, J. P. Novak, P. M. Campbell, D. Park, *Appl. Phys. Lett.* **82**, 2145 (2003).
16. P. Avouris, Z. Chen, V. Perebeinos, *Nat. Nanotechnol.* **2**, 605 (2007).
17. P. G. Collins, M. S. Arnold, P. Avouris, *Science* **292**, 706 (2001).
18. Materials and methods are available as supporting material on *Science* Online
19. Surface roughness measured by tapping-mode AFM was calculated over at least five different areas (1 μm by 1 μm) on the samples.
20. J. Liu *et al.*, *Chem. Phys. Lett.* **303**, 125 (1999).
21. S. Auvray *et al.*, *Nano Lett.* **5**, 451 (2005).
22. D. Chattopadhyay, L. Galeska, F. A. Papadimitrakopoulos, *J. Am. Chem. Soc.* **125**, 3370 (2003).
23. Although it has been shown by simulation only [see, for example, (*22*)] that NH_2 may selectively interact with neutral metallic nanotubes more favorably than with neutral semiconducting nanotubes, this trend is reversed by oxidation (such as at defect locations), as has been verified experimentally (*22*, *24*). Because all SWNT systems will have some level of defects [and as evidenced by the D′ band in micro-Raman spectra (figs. S9 and S11)], it is reasonable to expect that NH_2 should interact more strongly with semiconducting SWNTs in our system.
24. J. Lu *et al.*, *Small* **3**, 1566 (2007).
25. A. Nish, J.-Y. Hwang, J. Doig, R. J. Nicholas, *Nat. Nanotechnol.* **2**, 640 (2007).
26. R. Sharma, C. Y. Lee, J. H. Choi, K. Chen, M. S. Strano, *Nano Lett.* **7**, 2693 (2007).
27. This is calculated as broad mobility that accounts for full electrode width. The actual number of nanotubes

interfacing with the electrodes was not counted for all of these devices.

28. M. S. Dresselhaus, G. Dresselhaus, A. Jorio, A. G. Souza Filho, R. Saito, *Carbon* **40**, 2043 (2002).
29. A. Jorio *et al*., *Phys. Rev. Lett.* **86**, 1118 (2001).
30. B. Zhao *et al*., *J. Nanosci. Nanotechnol.* **4**, 995 (2004).
31. S. M. Keogh *et al*., *J. Phys. Chem. B* **108**, 6233 (2004).

32. This research was partially supported by the NSF-sponsored Center for Polymer Interface and Macromolecular Assemblies (CPIMA), Samsung Advanced Research Institute of Technology, Intel, and Applied Materials. M.C.L. acknowledges support from the Intelligence Community Postdoctoral Fellowship Program. M.R. acknowledges support from the NASA Graduate Student Research Fellowship. Z.B. acknowledges support from the Sloan Research Fellowship and Finmeccanica Faculty Scholar Award. We thank S. Hellstrom, H. W. Lee, J. N. Kurtin (Intel), and B. Chen (NASA Ames) for technical discussions.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/101/DC1
Materials and Methods
Figs. S1 to S11
Table S1
References

18 February 2008; accepted 29 May 2008
10.1126/science.1156588

Relativistic Spin Precession in the Double Pulsar

Rene P. Breton,[1]* Victoria M. Kaspi,[1] Michael Kramer,[2] Maura A. McLaughlin,[3,4] Maxim Lyutikov,[5] Scott M. Ransom,[6] Ingrid H. Stairs,[7] Robert D. Ferdman,[7,8] Fernando Camilo,[9] Andrea Possenti[10]

The double pulsar PSR J0737−3039A/B consists of two neutron stars in a highly relativistic orbit that displays a roughly 30-second eclipse when pulsar A passes behind pulsar B. Describing this eclipse of pulsar A as due to absorption occurring in the magnetosphere of pulsar B, we successfully used a simple geometric model to characterize the observed changing eclipse morphology and to measure the relativistic precession of pulsar B's spin axis around the total orbital angular momentum. This provides a test of general relativity and alternative theories of gravity in the strong-field regime. Our measured relativistic spin precession rate of $4.77^{\circ}{}^{+0^{\circ}.66}_{-0^{\circ}.65}$ per year (68% confidence level) is consistent with that predicted by general relativity within an uncertainty of 13%.

Spin is a fundamental property of most astrophysical bodies, making the study of its gravitational interaction an important challenge (*1*). Spin interaction manifests itself in different forms. For instance, we expect the spin of a compact rotating body in a binary system with another compact companion to couple gravitationally with the orbital angular momentum (relativistic spin-orbit coupling) and also with the spin of this companion (relativistic spin-spin coupling) (*2*, *3*). Observing such phenomena provides important tests for theories of gravity, because every successful theory must be able to describe the couplings and to predict their observational consequences. In a binary system consisting of compact objects such as neutron stars, one can generally consider the spin-orbit contribution acting on each body to dominate greatly the spin-spin contribution. This interaction results in a precession of the bodies' spin axis around the orbital angular momentum of the system, behavior we refer to as relativistic spin precession.

Although relativistic spin precession is well studied theoretically in general relativity (GR), the same is not true of alternative theories of gravity, and hence quantitative predictions of deviations from GR spin precession do not yet exist (*4*). For instance, it is expected that in alternative theories relativistic spin precession may depend on strong self-gravitational effects; that is, the actual precession may depend on the structure of a gravitating body (*4*). In the weak gravitational fields encountered in the solar system, these strong-field effects generally cannot be detected (*5*–*7*). Measurements in the strong-field regime near massive and compact bodies such as neutron stars and black holes are required. Relativistic spin precession has been observed in some binary pulsars [e.g., (*8*–*10*)], but it has usually only provided a qualitative confirmation of the effect. Recently, the binary pulsar PSR B1534+12 allowed the first quantitative measurement of this effect in a strong field, and although the spin precession rate was measured to low precision, it was consistent with the predictions of GR (*11*).

Here, we report a precision measurement of relativistic spin precession using eclipses observed in the double pulsar (*12*, *13*). This measurement, combined with observational access to both pulsar orbits in this system, allows us to constrain quantitatively relativistic spin precession in the strong-field regime within a general class of gravitational theories that includes GR.

PSR J0737−3039A/B consists of two neutron stars, both visible as radio pulsars, in a relativistic 2.45-hour orbit (*12*, *13*). High-precision timing of the pulsars, having spin periods of 23 ms and 2.8 s (hereafter called pulsars A and B, respectively), has already proven to be the most stringent test bed for GR in the strong-field regime (*14*) and enables four independent timing tests of gravity, more than any other binary system.

The orbital inclination of the double pulsar system is such that we observe the system almost perfectly edge-on. This coincidence causes pulsar A to be eclipsed by pulsar B at pulsar A's superior conjunction (*13*). The modestly frequency-dependent eclipse duration, about 30 s, corresponds to a region extending ~1.5 × 10^7 m (*15*). The light curve of pulsar A during its eclipse shows flux modulations that are spaced by half or integer numbers of pulsar B's rotational period (*16*). This indicates that the material responsible for the eclipse corotates with pulsar B. The relative orbital motions of the two pulsars and the rotation of pulsar B thus allow a probe of different regions of pulsar B's magnetosphere in a plane containing the line of sight and the orbital motion.

Synchrotron resonance with relativistic electrons is the most likely mechanism for efficient absorption of radio emission over a wide range of frequencies. In the model proposed by Lyutikov and Thompson (*17*), this absorbing plasma corotates with pulsar B and is confined within the closed field lines of a magnetic dipole truncated by the relativistic wind of pulsar A. The dipole magnetic moment vector makes an angle α with respect to the spin axis of pulsar B, whose orientation in space can be described by two angles: the colatitude of the spin axis with respect to the total angular momentum of the system, θ, and the longitude of the spin axis, ϕ (see Fig. 1 for an illustration of the system geometry). Additional parameters characterizing the plasma opacity, μ; the truncation radius of the magnetosphere, R_{mag}; and the relative position of pulsar A with respect to the projected magnetosphere of pulsar B, z_0, are also included in the model (*17*).

We monitored the double pulsar from December 2003 to November 2007 with the Green Bank Telescope in West Virginia; most of the data were acquired as part of the timing ob-

[1] Department of Physics, McGill University, Montreal, QC H3A 2T8, Canada. [2] Jodrell Bank Observatory, University of Manchester, Manchester M13 9PL, UK. [3] Department of Physics, West Virginia University, Morgantown, WV 26506, USA. [4] National Radio Astronomy Observatory, Green Bank, WV 24944, USA. [5] Department of Physics, Purdue University, West Lafayette, IN 47907, USA [6] National Radio Astronomy Observatory, Charlottesville, VA 22903, USA. [7] Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. [8] Laboratoire de Physique et Chimie de l'Environnement–CNRS, F-45071 Orleans cedex 2, France. [9] Columbia Astrophysics Laboratory, Columbia University, New York, NY 10027, USA. [10] Osservatorio Astronomico di Cagliari, Istituto Nazionale di Astrofisica, Poggio dei Pini, 09012 Capoterra, Italy.

*To whom correspondence should be addressed. E-mail: bretonr@physics.mcgill.ca

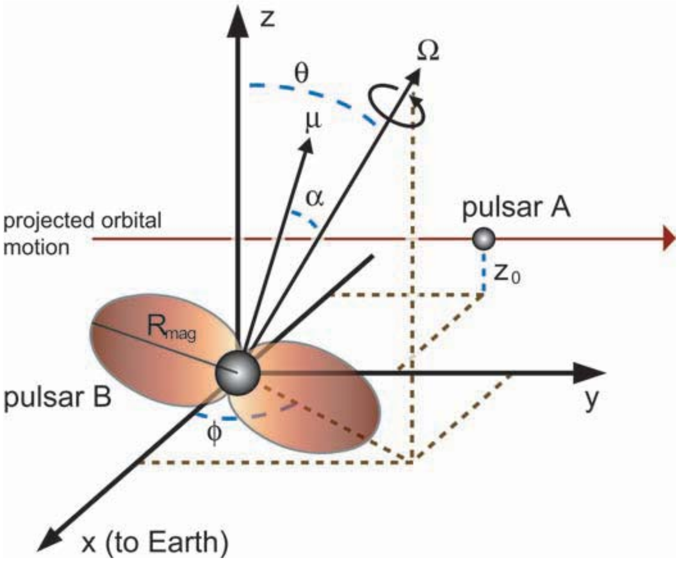


Fig. 1. Schematic view of the double pulsar system showing the important parameters for the modeling of pulsar A's eclipse (dimensions and angles are not to scale). Pulsar B is located at the origin of the cartesian coordinate system, whereas the projected orbital motion of pulsar A during its eclipse is parallel to the *y* axis at a constant z_0 as seen from Earth, which is located toward the positive *x* axis. Because the orbital inclination is almost perfectly edge-on (*14*), we can approximate the *z* axis to be coincident with the orbital angular momentum. The spin axis of pulsar B, whose spatial orientation is described by θ and ϕ, is represented by the Ω vector. The magnetic axis of pulsar B corresponds to the μ vector and makes an angle α with respect to Ω. Lastly, the absorbing region of the dipolar magnetosphere of pulsar B, truncated at radius R_{mag}, is shown as a shaded red region.

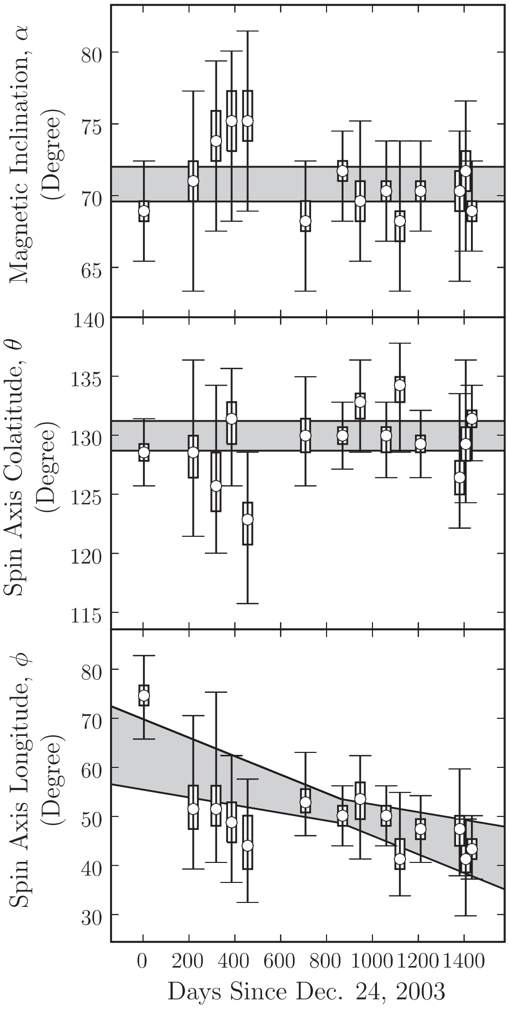


Fig. 2. Evolution of pulsar B's geometry as a function of time. The marginalized posterior probability distribution of the magnetic inclination (α), the colatitude of the spin axis (θ), and the longitude of the spin axis (ϕ) of pulsar B are shown from top to bottom, respectively. For each data point, the circle represents the median value of the posterior probability density, whereas the box and the bar indicate the 1σ and 3σ confidence intervals, respectively. The gray regions are the 3σ confidence regions derived from the joint time-dependent model fitting. For clarity, multiple eclipses are displayed as single data points when observed over an interval of about a week.

servations reported in (*14*). The data used for our analysis were taken at 820 MHz with the SPIGOT instrument (*18*), which provides 1024 frequency channels across a 50-MHz bandwidth. Data for a total of 63 eclipses of pulsar A were collected over the 4-year period, with many obtained during semi-annual concentrated observing campaigns. We dedispersed each eclipse data set by adding time shifts to frequency channels in order to compensate for the frequency-dependent travel time of radio waves in the ionized interstellar medium, and we then folded them at the predicted spin period of pulsar A by using the pulsar analysis packages PRESTO (*19*) and SIGPROC (*20*) [see (*14*) for details about the radio timing]. Next, we extracted the relative pulsed flux density of pulsar A by fitting each folded interval for the amplitude of a high signal-to-noise ratio pulse profile template made from the integrated pulse observed during the several-hour observation that includes each eclipse. Lastly, we normalized the flux densities so the average level outside the eclipse region corresponded to unity. We chose the time resolution of our eclipse light curves to equal, on average, four individual pulses of pulsar A (~91 ms).

In addition to the flux density, we determined the orbital phase and the spin phase of pulsar B corresponding to each data point of our time series. Orbital phases were derived from the ephemeris published in (*14*). Spin phases were empirically measured from data folded at the predicted period of pulsar B in a way similar to that described above for pulsar A. Over the 4-year monitoring campaign, we found notable changes in pulsar B's pulse profile, likely due to the precession of its spin axis, which were also reported in (*21*). Around 2003, the average pulse profile was unimodal, resembling a Gaussian function. It evolved such that, by 2007, it displayed two narrow peaks. Using the pulse peak maximum as a fiducial reference point is certainly not appropriate. We find, however, that the unimodal profile gradually became wider and then started to form a gap near the center of its peak. Since then, the outer edges of the pulse profile have not significantly changed, but the gap evolved such that two peaks are now visible. This lets us presume that the underlying average profile is reminiscent of a Gaussian-like profile to which some "absorption" feature has been superimposed near the center, leaving a narrow peak on each side. We therefore defined the fiducial reference point to lie at the center of the unimodal "envelope" that we reconstructed from the first 10 Fourier bins of the pulse profile, which contains 512 bins in total (see fig. S2 for an illustration of the pulse profile evolution).

We implemented the eclipse modeling of our data in two steps: the fitting of individual eclipse profiles and the search for evolution of the geometry of pulsar B. We first searched the full phase space to identify best-fit values of six parameters [see supporting online material (SOM) for more details]. Then, we reduced the number

Table 1. Geometrical parameters of pulsar B derived from the eclipse model fitting. The presented values include priors related to systematic uncertainties. The epoch of $\phi = \phi_0$ is 2 May 2006 (MJD 53857).

Parameter	Mean	Median	68.2% confidence	99.7% confidence
α_0	70.92°	70.94°	(70.49°, 71.31°)	(69.68°, 72.13°)
θ_0	130.02°	130.02°	(129.58°, 130.44°)	(128.79°, 131.37°)
ϕ_0	51.21°	51.20°	(50.39°, 52.03°)	(48.80°, 53.72°)
Ω_B	4.77° year^{-1}	4.76° year^{-1}	(4.12°, 5.43°) year^{-1}	(2.89°, 6.90°) year^{-1}

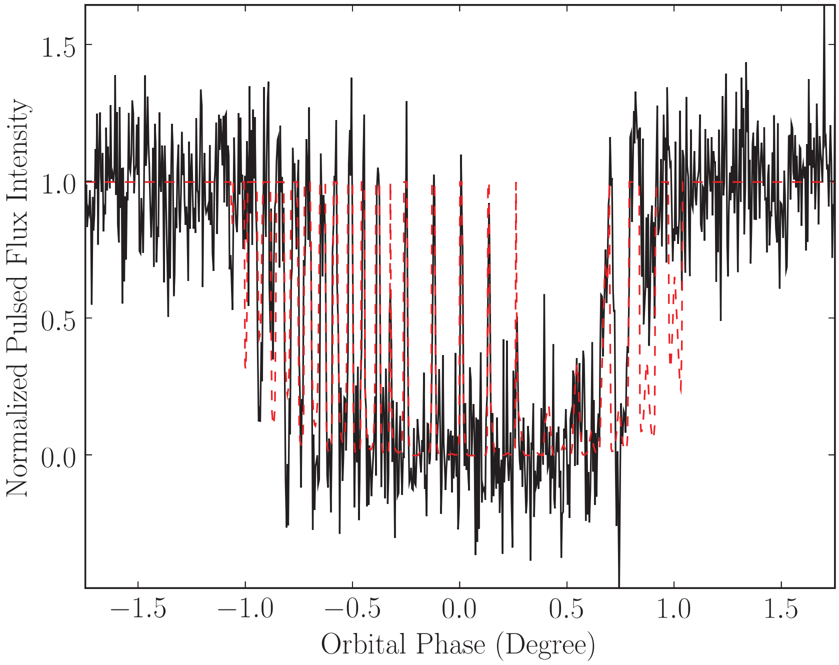


Fig. 3. Average eclipse profile of pulsar A consisting of eight eclipses observed at 820 MHz over a 5-day period around 11 April 2007 (black line) along with a model eclipse profile (red dashed line). The relative pulsed flux density of pulsar A is normalized so that the average level outside the eclipse region is unity. The resolution of each data point is ~91 ms, whereas 1° in orbital phase corresponds to 24.5 s. Near orbital phase 0.0, the spikes are separated by the spin period of pulsar B.

of free parameters to the subset (θ, ϕ, and α) describing the orientation of pulsar B's spin and magnetic axes by fixing the other parameters to their best-fit values: $\mu = 2$, $R_{\mathrm{mag}} = 1.29°$ (projected value in terms of orbital phase), and $z_0/R_{\mathrm{mag}} = -0.543$ (Fig. 1). Lastly, we performed a high-resolution mapping of the likelihood of this subspace in order to investigate subtle changes in the geometry. Lyutikov and Thompson (*17*) predicted that such changes, because of relativistic spin precession, could affect the eclipse light curve. In principle, relativistic spin precession of pulsar B's spin axis around the total angular momentum should induce a secular change of the longitude of the spin axis, ϕ, whereas the magnetic inclination, α, and the colatitude of the spin axis, θ, are expected to remain fixed over time. Indeed, from model fitting, we find no significant time evolution of α and θ, whereas ϕ does change. Because of correlation between the parameters, we jointly evaluated the best-fit geometry of pulsar B by using a time-dependent model in which $\alpha = \alpha_0$ and $\theta = \theta_0$ are constants and ϕ varies linearly with time; i.e., $\phi = \phi_0 - \Omega_B t$, where Ω_B is the rate of change of pulsar B's spin axis longitude and the epoch of $\phi = \phi_0$ is 2 May 2006 [Mean Julian Day (MJD) 53857]. Figure 2 shows the time evolution of the parameters and the fit derived from this joint time-dependent model (Table 1). The precession rate Ω_B of $4.77°^{+0°.66}_{-0°.65}$ year^{-1} (*22*) agrees with the precession rate predicted by GR (*23*), 5.0734° ± 0.0007° year^{-1} (*24*), within an uncertainty of 13% (68% confidence level).

This relatively simple model (*17*) is able to reproduce the complex phenomenology of the eclipses (Fig. 3 and movie S1) except at the eclipse boundaries, where slight magnetospheric distortions or variations in plasma density are likely to occur. Fits including the egress generally are poor in the central region where we observe narrow modulation features, which are critical for determining pulsar B's geometry. For this reason, we excluded the egress from the fits, using orbital phases between −1.0° and 0.75° (Fig. 3). We accounted for systematics introduced by the choice of the region to fit in the priors of our Bayesian model (SOM). This improved the fit of the model throughout the center region of the eclipse while still producing qualitatively good predictions near the eclipse egress. The overall success of the model implies that the geometry of pulsar B's magnetosphere is accurately described as predominantly dipolar; a pure quadrupole, for instance, does not reproduce the observed light curves. Although the model does not exclude the possibility that higher-order multipole components may exist close to the surface of pulsar B, our modeling supports the conclusions (*17*) that these eclipses yield direct empirical evidence supporting the long-standing assumption that pulsars have mainly dipolar magnetic fields far from their surface.

The direct outcome from modeling the eclipse profile evolution is a measurement of the effect of relativistic spin precession (see movie S2 for an illustration of the time evolution of the eclipse). We can use the inferred precession rate to test GR (Fig. 4) and to further constrain alternative theories of gravity and the strong-field aspects of relativistic spin precession. We use the generic class of relativistic theories that are fully conservative (Lorentz-invariant) and based on a Lagrangian, as introduced by Damour and Taylor (*4*). In this way, we can study the constraints of our observations on theories of gravity by describing the spin-orbit interaction within a specific theory that couples functions appearing in the corresponding part of the Lagrangian. In this framework, we can write the precession rate of pulsar B in a general form, $\Omega_B = \sigma_B L/a^3_R(1 - e^2)^{3/2}$, where L is the orbital angular momentum of the system, a_R is the semimajor axis of the relative orbit between the pulsars, e the eccentricity of the orbit, and σ_B is a generic strong-field spin-orbit coupling constant. Because L and a_R are not directly measurable, it is more convenient to write the above expression with use of observable Keplerian and post-Keplerian parameters. Although alternative forms generally involve a mixture of gravitational theory–dependent terms, the particular choice $\Omega_B = \frac{x_A x_B}{s^2} \times \frac{n^3}{1-e^2} \times \frac{c^2 \sigma_B}{G}$ is the only one that does not incorporate further theoretical terms other than the spin-orbit coupling constant, σ_B; the speed of light, c; and a generalized gravitational constant for the interaction between the two pulsars, G. In this expression, the Keplerian parameters e and $n = 2\pi/P_b$, the angular orbital frequency, are easily measurable for any binary system. On the other hand, the post-Keplerian Shapiro delay shape parameter, s, equivalent to the sine of the orbital inclination angle (*4*), requires relatively edge-on orbits to be observed. Measurement of the projected semi-major axes of the two orbits (*25*), x_A and x_B, found in the above equation necessitates that each body must be able to be timed. Therefore, the double pulsar is the only relativistic binary system that allows a direct constraint on the spin-orbit coupling in general theories of gravity. By using the inferred preces-

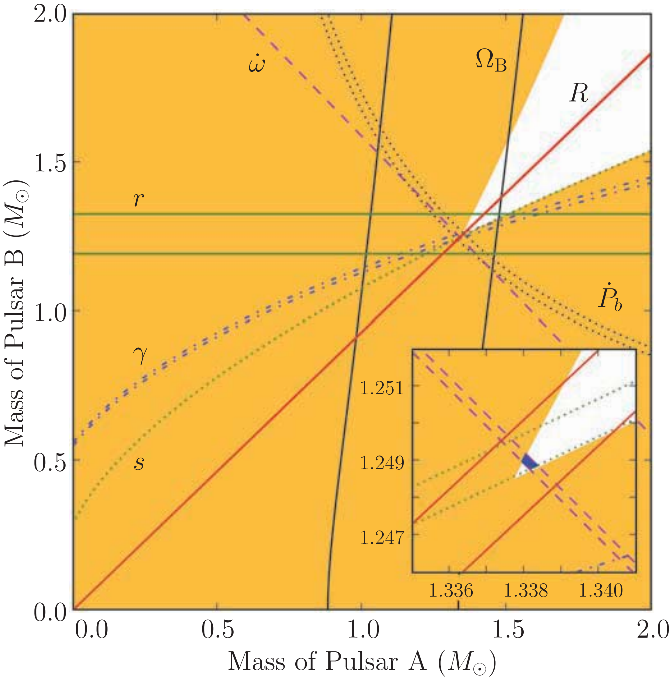


Fig. 4. Mass-mass diagram illustrating the present tests constraining general relativity in the double pulsar system. (**Inset**) An expanded view of the region where the lines intersect. If general relativity is the correct theory of gravity, all lines should intersect at common values of masses. The mass ratio ($R = x_B/x_A$) and five post-Keplerian parameters (s and r, Shapiro delay shape and range; $\dot{\omega}$, periastron advance; $\dot{P}_b$, orbital period decay due to the emission of gravitational waves; and γ, gravitational redshift and time dilation) were reported in (*14*). Shaded orange regions are unphysical solutions because $\sin i \leq 1$, where i is the orbital inclination. In addition to allowing a test of the strong-field parameter $\left(\frac{c^2\sigma_B}{G}\right)$, the spin precession rate of pulsar B, Ω_B, yields a new constraint on the mass-mass diagram. $M_\odot$ is the mass of the Sun.

sion rate of $\Omega_B = 4.77^{\circ}{}^{+0^{\circ}.66}_{-0^{\circ}.65}$ year^{-1}, we derive $\left(\frac{c^2\sigma_B}{G}\right) = 3.38^{+0.49}_{-0.46}$. Every successful theory of gravity in the given generic framework must predict this value: These observations provide a strong-field test of gravity that complements and goes beyond the weak-field tests of relativistic spin precession (*26*). In GR, we expect to measure $\left(\frac{c^2\sigma_B}{G}\right)_{\rm GR} = 2 + \frac{3}{2}\frac{m_A}{m_B} = 3.60677 \pm 0.00035$, where we have used the masses determined from the precisely observed orbital precession and the Shapiro delay shape parameter under the assumption that GR is correct (*14*). Comparing the observed value with GR's predictions, we find $\left(\frac{c^2\sigma_B}{G}\right)_{\rm obs}/\left(\frac{c^2\sigma_B}{G}\right)_{\rm GR} = 0.94 \pm 0.13$. Hence, GR passes this test of relativistic spin precession in a strong-field regime, confirming, within uncertainties, GR's effacement property of gravity even for spinning bodies, that is, the notion that strong internal gravitational fields do not prevent a compact rotating body from behaving just like a spinning test particle in an external weak field (*27*).

The spin precession rate, as well as the timing parameters entering in the calculation of $\left(\frac{c^2\sigma_B}{G}\right)$, are all independent of the assumed theory of gravity. If the main contribution limiting the precision of this new strong-field test comes from the inferred spin precession rate, we expect that the statistical uncertainty should decrease significantly with time, roughly as the square of the monitoring baseline for similar quantity and quality of eclipse data. The contribution of systematics to the error budget should also decrease, but its functional time dependence is difficult to estimate. Although the orbital and spin phases of pulsar B are input variables to the eclipse model, our ability to determine the orientation of pulsar B in space does not require the degree of high-precision timing needed for measurement of post-Keplerian parameters; evaluating spin phases to the percent level, for instance, is sufficient. Therefore, the intrinsic correctness of the model and its ability to reproduce future changes in the eclipse profile because of evolution of the geometry are the most likely limitations to improving the quality of this test of gravity, at least until the measured precession rate reaches a precision comparable with the timing parameters involved in the calculation of $\left(\frac{c^2\sigma_B}{G}\right)$. Better eclipse modeling could be achieved from more sensitive observations, and thus new-generation radio telescopes such as the proposed Square Kilometer Array could help make important progress. Pulsar A does not show evidence of precession (*28*, *29*) likely because its spin axis is aligned with the orbital angular momentum; it should therefore always remain visible, thus allowing long-term monitoring of its eclipses. Pulsar B, however, could disappear if spin precession causes its radio beam to miss our line of sight (*21*). In this event, we would need to find a way to circumvent the lack of observable spin phases for pulsar B, which are necessary to the eclipse fitting.

References and Notes

1. C. Will, *Living Rev. Relativity* **4**, 4 (2001).
2. The contribution from a classical quadrupolar moment is negligible for compact bodies.
3. R. F. O'Connell, in *Experimental Gravitation: Proceedings of Course 56 of the International School of Physics "Enrico Fermi,"* B. Bertotti, Ed. (Academic Press, New York, 1974), p. 496.
4. T. Damour, J. H. Taylor, *Phys. Rev. D* **45**, 1840 (1992).
5. T. Damour, G. Esposito-Farèse, *Phys. Rev. D* **46**, 4128 (1992).
6. T. Damour, G. Esposito-Farèse, *Class. Quantum Gravity* **9**, 2093 (1992).
7. T. Damour, G. Esposito-Farèse, *Phys. Rev. D* **53**, 5541 (1996).
8. J. M. Weisberg, R. W. Romani, J. H. Taylor, *Astrophys. J.* **347**, 1030 (1989).
9. M. Kramer, *Astrophys. J.* **509**, 856 (1998).
10. A. W. Hotan, M. Bailes, S. M. Ord, *Astrophys. J.* **624**, 906 (2005).
11. I. H. Stairs, S. E. Thorsett, Z. Arzoumanian, *Phys. Rev. Lett.* **93**, 141101 (2004).
12. M. Burgay *et al.*, *Nature* **426**, 531 (2003).
13. A. G. Lyne *et al.*, *Science* **303**, 1153 (2004); published online 8 January 2004 (10.1126/science.1094645).
14. M. Kramer *et al.*, *Science* **314**, 97 (2006); published online 14 September 2006 (10.1126/science.1132305).
15. V. M. Kaspi *et al.*, *Astrophys. J.* **613**, L137 (2004).
16. M. A. McLaughlin *et al.*, *Astrophys. J.* **616**, L131 (2004).
17. M. Lyutikov, C. Thompson, *Astrophys. J.* **634**, 1223 (2005).
18. D. L. Kaplan *et al.*, *Publ. Astron. Soc. Pacific* **117**, 643 (2005).
19. S. M. Ransom, S. S. Eikenberry, J. Middleditch, *Astrophys. J.* **124**, 1788 (2002). PRESTO is freely available at www.cv.nrao.edu/~sransom/presto/.
20. D. R. Lorimer. SIGPROC is freely available at http://sigproc.sourceforge.net.
21. M. Burgay *et al.*, *Astrophys. J.* **624**, L113 (2005).
22. Unless otherwise stated, uncertainties are quoted at the 68% confidence level.
23. B. M. Barker, R. F. O'Connell, *Phys. Rev. D* **12**, 329 (1975).
24. The uncertainty on the predicted GR spin precession rate arises because the value depends on the masses of the system, which are determined from two measured post-Keplerian parameters: the Shapiro delay s parameter and the advance of periastron, $\dot{\omega}$.
25. The projected semi-major axes are expressed in terms of light travel time across the orbit.
26. R. F. O'Connell, http://arxiv.org/abs/0804.3806 (2008).
27. T. Damour in *Three Hundred Years of Gravitation*, S. W. Hawking, W. Israel, Eds. (Cambridge Univ. Press, Cambridge, 1987), pp. 128–198.
28. R. N. Manchester *et al.*, *Astrophys. J.* **621**, L49 (2005).
29. R. D. Ferdman *et al.*, *40 Years of Pulsars: Millisecond Pulsars, Magnetars and More*, vol. 983 of *American Institute of Physics Conference Series*, C. Bassa, Z. Wang, A. Cumming, V. M. Kaspi, Eds. (American Institute of Physics, Melville, NY, 2008), pp. 474–478.
30. We thank N. Wex, T. Damour, R. F. O'Connell, R. D. Blandford, and W. G. Unruh for stimulating discussions. The National Radio Astronomy Observatory is a facility of the NSF operated under cooperative agreement by Associated Universities, Incorporated. This work was supported in part by the Natural Sciences and Engineering Research Council Discovery Grant program, the Canada Foundation for Innovation, the Canadian Institute for Advanced Research, an FQRNT (Le Fonds Québécois de la Recherche sur la Nature et les Technologies) team grant, the Canada Research Chair Program, the McGill University Lorne Trottier Chair in Astrophysics and Cosmology, and the Ministero dell'Istruzione dell'Università e della Ricerca under the grant PRIN-2005024090.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/104/DC1
Materials and Methods
Figs. S1 and S2
References
Movies S1 and S2

18 April 2008; accepted 4 June 2008
10.1126/science.1159295

Characterization of Step-Edge Barriers in Organic Thin-Film Growth

Gregor Hlawacek,[1] Peter Puschnig,[2] Paul Frank,[3] Adolf Winkler,[3] Claudia Ambrosch-Draxl,[2] Christian Teichert[1]*

Detailed understanding of growth mechanisms in organic thin-film deposition is crucial for tailoring growth morphologies, which in turn determine the physical properties of the resulting films. For growth of the rodlike molecule *para*-sexiphenyl, the evolution of terraced mounds is observed by atomic force microscopy. Using methods established in inorganic epitaxy, we demonstrate the existence of an additional barrier (0.67 electron volt) for step-edge crossing—the Ehrlich-Schwoebel barrier. This result was confirmed by transition state theory, which revealed a bending of the molecule at the step edge. A gradual reduction of this barrier in the first layers led to an almost layer-by-layer growth during early deposition stage. The reported phenomena are a direct consequence of the complexity of the molecular building blocks versus atomic systems.

Although organic electronics (*1*) with devices like organic thin-film transistors and organic light-emitting diodes have started to enter the consumer market, there is still—in contrast to inorganic epitaxy—a severe lack of understanding of the basic growth mechanisms. In inorganic epitaxy—mainly of metal films—growth kinetics can substantially determine growth morphologies. In particular, interlayer mass transport is necessary to allow for the layer-by-layer growth mode (*2*) needed to avoid formation of rough surfaces. Field ion microscopy has revealed that, under certain growth conditions, diffusing adatoms are reflected at descending step edges (*3*). These dynamics have been explained by the presence of an additional activation barrier for an adatom when crossing a step edge (*3*–*5*), which is usually called the Ehrlich-Schwoebel barrier (ESB). For an active ESB, terraced growth mounds form by repeated two-dimensional nucleation of islands of monoatomic thickness (*6*). In the complete absence of interlayer mass transport, the surface fractions in the individual layers follow a Poisson distribution (*7*). Moreover, deep trenches form between the mounds that ideally never quite close, and their formation has been referred to as the Zeno effect (*8*). Interpretation of this effect is based on the presence of a step-edge barrier and allows predictions on quantitative roughness parameters and on the cross-sectional shape of the resulting mounds (*8*). With the advent of modern scanning probe techniques, this shape can be analyzed in detail and allows for a determination of the ESB (*9*–*11*).

In organic thin-film growth, a wide variety of growth morphologies is observed because of the complexity of the building blocks and their interactions. Recently, the formation of mounds in organic thin-film growth (*12*–*14*), as well as theoretical efforts to determine the ESB (*15*), has been reported. In contrast to metal or inorganic semiconductor epitaxy, in which single atoms diffuse, a richer spectrum of growth mechanisms is expected in organic film growth because molecules are anisotropic and have internal degrees of freedom that open novel diffusion channels accompanied by changes in the molecular conformation.

We report on the growth processes involved in the formation of *para*-sexiphenyl (6P) mounds on mica surfaces modified by ion bombardment so that the molecules pack upright on the surface (*16*). By means of atomic force microscopy (AFM), we evaluate the thickness-dependent evolution of the overall surface roughness, as well as the mound shapes. Using a model developed for inorganic epitaxy, we determined the step-edge barrier for interlayer diffusion of 6P. We emphasize that a diffusion scenario applicable to organic semiconductors differs substantially from classical diffusion mechanisms. Our experiments are supported by transition state theory calculations and molecular dynamics simulations that quantitatively confirm the measured barrier height and illuminate the underlying complex process. Not only do we observe an ESB with an energy of several tenths of an electron volt, we also show its layer dependence.

6P, a blue-light–emitting, π-conjugated molecule, has six phenyl rings connected by single bonds, and van der Waals dimensions of 2.85 nm by 0.35 nm by 0.67 nm. For the film preparation, freshly cleaved mica was inserted into the ultrahigh-vacuum system (base pressure 2 × 10^{-10} mbar) and bombarded with low-energy Ar^+ ions to destroy the ordered structure of the surface layer. In contrast, on the freshly cleaved mica, sizable molecule-substrate interactions would lead to a wetting layer and induce the growth of needle-like structures composed of lying molecules (*17*–*19*). Sputtering was stopped as soon as low-energy electron diffraction showed no ordered structure (*16*). Subsequently, 6P was grown in the absence of a wetting layer (*20*) by means of physical vapor deposition from a homemade Knudsen cell with a growth rate *F* of 0.3 nm/min at room temperature.

In Fig. 1, A to C, we show AFM images of three 6P films with growing thickness (namely 4, 10, and 30 nm) that demonstrate the formation of mounds with an increasing number of terraces. The cross section of such a single terraced mound, as depicted in the inset of Fig. 1B, reveals that the mounds are characterized by 2.6 ± 0.3 nm–high steps. This value corresponds to the height of a layer formed by nearly upright-standing molecules, as found in the bulk structure (*21*). As thickness increased, the lateral mound shape changed from ill-defined to roughly polygonal. However, no orientation relation was observed between individual

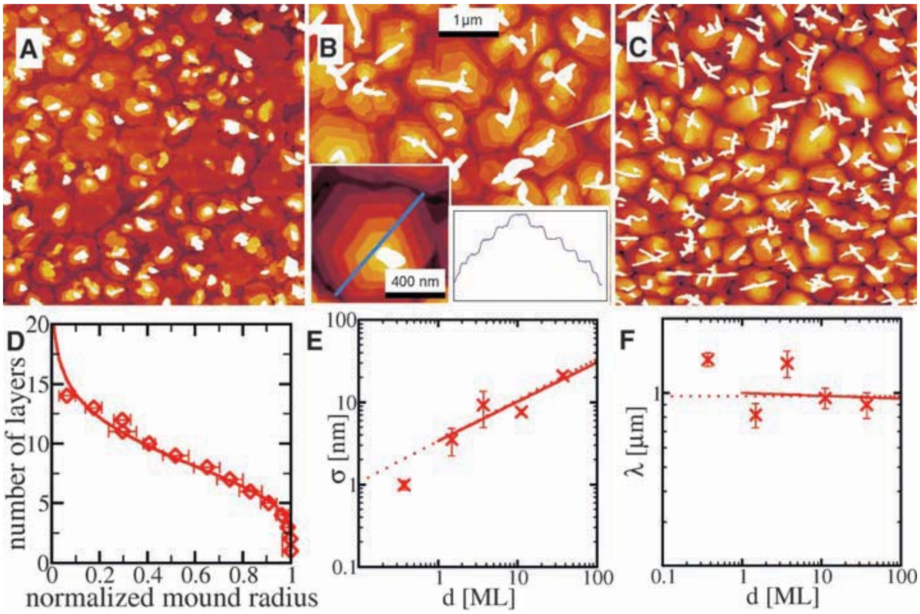


Fig. 1. (**A** to **C**) Morphological analysis of 6P thin films (*z* scale: 20, 35, and 50 nm, respectively). The inset in (B) shows a single mound and the corresponding cross section. (**D**) Normalized averaged cross section of five mounds found in the 30-nm-thick film. The solid line represents the cumulative Poisson distribution used to fit the data. (**E**) Evolution of the RMS roughness σ and (**F**) of the average mound separation λ with film thickness. The error bars correspond to the statistical uncertainty in the data measured by AFM.

[1]Institute of Physics, University of Leoben, 8700 Leoben, Austria. [2]Chair of Atomistic Modelling and Design of Materials, University of Leoben, 8700 Leoben, Austria. [3]Institute of Solid State Physics, Graz University of Technology, 8010 Graz, Austria.

*To whom correspondence should be addressed. E-mail: teichert@unileoben.ac.at

mounds. The single mounds are separated by deep crevices that do not close with ongoing deposition. As more material was deposited, branched needle-like structures formed on top of the mounds. This morphology is identified as the typical needles that 6P forms on ordered substrates and consists of flat-lying molecules (*18*, *19*).

To analyze the cross-sectional shape in detail, we used a method introduced by Kalff *et al.* (*9*) to calculate an average mound profile from cross sections. The normalized mound shape, presented in Fig. 1D, also shows the steep bottom section forming the trenches and the flat top, leading to a nonconstant slope of the mound. The mound profile can be fitted by a cumulative Poisson distribution (*7*) (solid line in Fig. 1D), in agreement with the theoretical description by Elkinani and Villain (*8*), who introduced the term "Zeno model" to reflect the frustration of film growth in the crevices between the mounds.

Two experimentally accessible quantities were used as a fingerprint for the existence of an ESB, namely, the root mean square (RMS) roughness σ, and the average mound separation λ. By applying scaling laws, we can quantify the evolution of the mounded growth front (*11*). Figure 1E shows the RMS roughness σ as a function of the film thickness. The evolution of σ is in agreement with the observation mentioned above that the mound profile follows a cumulative Poisson distribution for which the standard deviation is given by the square root of the mean value, the film thickness *d* in this case.

This result allows one to predict that for mounds with the described shape, the RMS roughness σ follows a $\sqrt{d}$ behavior, so that the growth exponent should be β = 0.5. A power-law fit to the data indeed reveals a growth exponent β = 0.49, as represented by the solid line in Fig. 1E. Moreover, we find that the evolution of the average mound separation λ—as obtained by fitting the height-height correlation function with a modified form suitable for mounded surfaces (*22*)—can be expressed by a power law with a dynamic growth exponent of 1/*z* = −0.01 (solid line in Fig. 1F). We note that 1/*z* = 0 (dashed line in Fig. 1F) means that the mound separation remains unchanged after the first-layer nucleation has stopped; all mass transport between individual mounds is suppressed, and no coalescence or ripening occurs. Such a behavior requires a barrier, the ESB (*3*–*5*), that prevents the molecules from diffusing to a lower terrace of the mounds and subsequently to the neighboring mound. This scenario is known as the Zeno model (*8*).

The height of the ESB E_{ES} governs the rate ν′ at which molecules diffuse downward across a terrace edge, i.e., the interlayer jump rate, which can be written as

$$\nu' = \nu e^{-\frac{E_{ES}}{k_B T}} \qquad (1)$$

Here, the first term $\nu \propto e^{-\frac{E_{\mathrm{D}}}{k_{\mathrm{B}}T}}$ is the jump rate for diffusion on a single terrace, which is controlled by the comparably small energy barrier for on-terrace diffusion E_{D}; the second term arises from the additional energy barrier E_{ES} at the step edge. Following a method successfully used for Pt/Pt(111) homoepitaxy (*10*, *9*), the interlayer jump rate ν′ can be estimated from the top terrace diameter *l* and the flux *F* (*10*):

$$\ell \propto \left(\frac{\nu'}{F}\right)^{\frac{1}{5}} \qquad (2)$$

To determine the height of the ESB from the equations above, we evaluate the on-terrace jump rate ν from kinetic nucleation theory (*23*). In the presence of an ESB, large enough to effectively suppress interlayer mass transport, the mound separation λ is related to the nucleation density *N* (*24*):

$$\frac{1}{\lambda^2} \propto N \propto \left(\frac{F}{\nu}\right)^{\frac{i^*}{D(i^*+2)}} \qquad (3)$$

because the initial seeds are determining the final density of the growth mounds, and no coarsening or smoothing occurs. For a two-dimensional diffusion process (*D* = 2), and a critical nucleus size of *i** = 1, the exponent is 1/6. The experimentally determined mound separation of 1 μm, the deposition flux *F*, and the measured top terrace diameter between 20 and 60 nm result in a value of 0.67 ± 0.04 eV for the ESB of upright-standing 6P.

To cross-check the experimental finding, we calculate the minimum-energy path for step-edge crossing by means of a molecular model. The activation barrier for interlayer mass transport is given by the saddle-point energy on the potential energy surface. The large size of the required simulation cell and the numerous structural degrees of freedom preclude a purely ab initio approach, so we used a combination of first-principles density functional calculations with an empirical force field method (*16*). The resulting energetics of the transition path, as well as snapshots of the molecule conformation, are shown in Fig. 2. The binding energy of a 6P molecule on top of an existing island of upright-standing molecules is computed to be −1.29 eV, where the molecule is lying almost flat on top of the 6P(001)

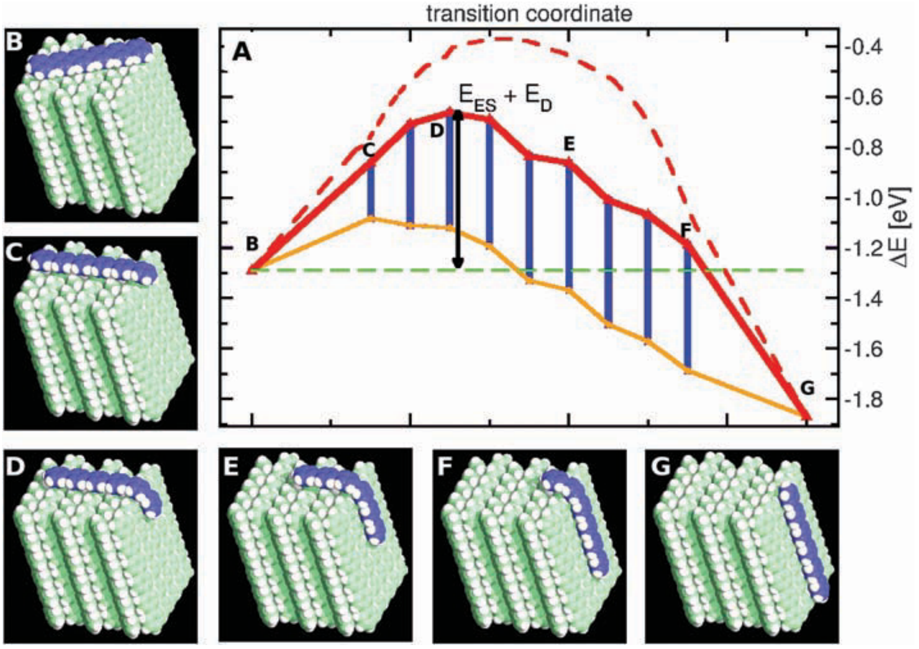


Fig. 2. (**A**) Energies associated with the transition path for diffusion of a 6P molecule over a step edge, resulting in a total activation barrier of 0.63 eV (indicated by the black arrow). The red solid line indicates the total binding energy of the additional molecule, which is composed of the attractive intermolecular interaction (orange) and the positive bending energy (blue bars). Labels B to G correspond to situations illustrated in images (B) to (G). The green dashed line denotes the energy in the initial position (B). (**B** to **G**) Snapshots of the transition path; (D) corresponds to the configuration at the transition state. For the red dashed line, rigid molecules have been assumed for the transition state calculation.

Table 1. Expected (*Z*eno model) and measured 6P coverage in the first four layers.

Layer	1-nm-thick film		4-nm-thick film	
	Expected (%)	Measured (%)	Expected (%)	Measured (%)
1	30.9	50.0 ± 2.0	75.5	94.5 ± 4.0
2	5.3	0.1 ± 0.01	41.8	84.2 ± 3.0
3	0.6	0.000	16.9	26.5 ± 1.0
4	0.005	0.000	4.6	9.3 ± 0.5

surface (Fig. 2B). To reach its global energetic minimum in which the 6P molecule is integrated into the (100) plane with a binding energy of −1.87 eV (Fig. 2G), it has to overcome a barrier of 0.63 eV. Taking into account the on-terrace diffusion barrier E_D of only 0.02 eV, also calculated by the same method, results in an ESB of 0.61 eV, in excellent agreement with the experimental value.

An analysis of the transition pathway reveals that the 6P molecule diffuses toward the [100] edge with its long axis perpendicular to the edge and gradually slides down the (100) plane by bending over the edge between (001) and (100) facets as illustrated in the snapshots (Fig. 2, B to G). At first glance this seems surprising. However, it is known that the 6P molecule can bend around defects or surface corrugations of the substrate (*25*). We can break down the contributions to the total energy of the attached 6P molecule into attractive intermolecular interactions and an additional bending energy that has to be spent. For instance, at the transition state (Fig. 2D), intermolecular interactions are only slightly less effective than they are for the top configuration (B), whereas more than 70% of the ESB is due to the energy cost related to bending the molecule. Using a rigid molecule for the transition state calculation leads to an ESB of 0.90 eV (red dashed line in Fig. 2A), which is distinctly larger than the value given above and also in disagreement with the experimental counterpart. Thus, breaking nearly all intermolecular bonds at the same time is energetically unfavorable compared to the bending scenario described above. Also downward funneling and push-out mechanisms can be excluded from our experimental analysis (*26*). Likewise, exchange mechanisms—common in metal epitaxy (*27*)—are not observed by the simulations.

Analyzing the experimental data for the early growth stage, we observe a peculiarity not seen in inorganic film growth. First, the distribution of the layer fractions for this growth stage (Table 1) reveals a substantial deviation from the Poisson distribution. The latter is expected in the absence of interlayer mass transport, which should be the case in the limit of a large ESB. Indeed, the growth front appears more layer-by-layer–like with less material in the upper layers. As pointed out by Krug and coworkers (*10*), the number fraction of first-layer islands *f* with size *L* that have nucleated a second layer island on top offers another independent possibility to quantify the ESB from the AFM data. Thus, we calculate the lower step-edge barrier in the first layer by looking at the number of islands that have nucleated in the second level

$$f = 1 - e^{-\left(\frac{L}{L_c}\right)^{7}} \qquad (4)$$

Here, L_c denoting the critical island size

$$L_c = \left(\frac{7\nu'}{2\alpha^4 FN}\right)^{\frac{1}{7}} \qquad (5)$$

again depends on the interlayer jump rate ν′. Further, $\alpha = A/C^2$ takes into account the ramified shape of the islands, with *A* being the island area and *C* the circumference, which is also accessible from the AFM data. Using the above relation, we can calculate ν′ and consequently the barrier height E_{ES} for the initial growth stage. With the data obtained from the 1-nm-thick film, we arrive at an ESB of 0.26 eV, which is appreciably lower than the value obtained for the mounded films. In metal homoepitaxy (*11*), such a decrease of the ESB is due to the noncompact island shape in the first layers allowing easy atom descent, e.g., at kink sites (*28*). Our second finding, however, points to a different mechanism active here.

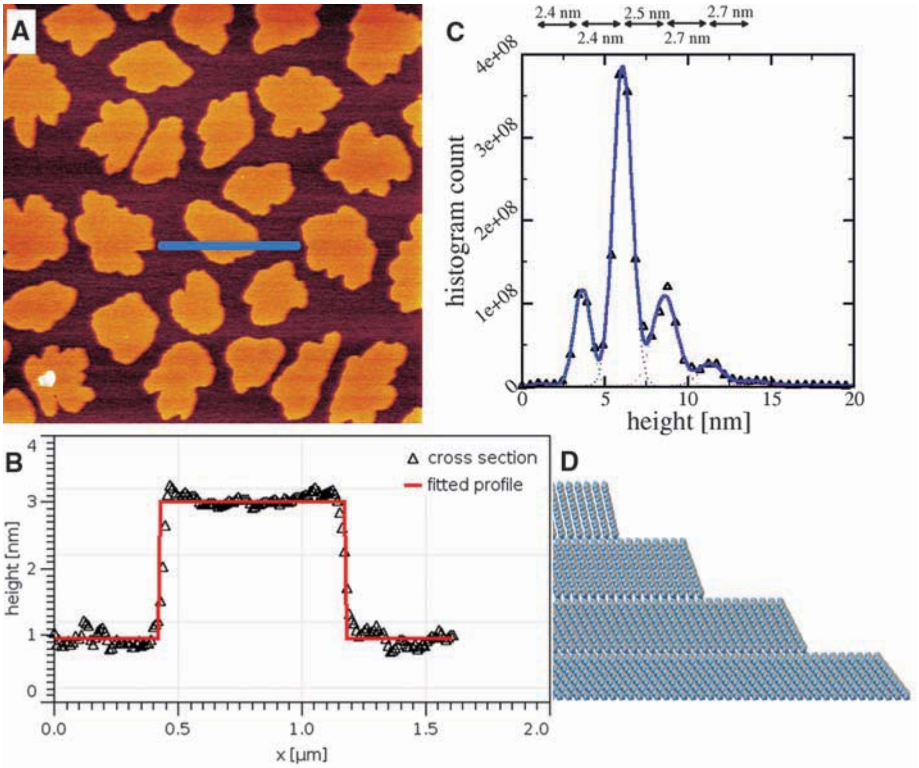


Fig. 3. Different heights for the first few layers. (**A**) AFM image of a 1-nm-thick 6P film and (**B**) marked cross section through one of the islands. The fitted height is 2.05 ± 0.1 nm. (**C**) Height analysis of the 4-nm-thick 6P film, revealing an increasing layer thickness for the first four layers. (**D**) Schematic view of the molecular tilt angle in the first four layers.

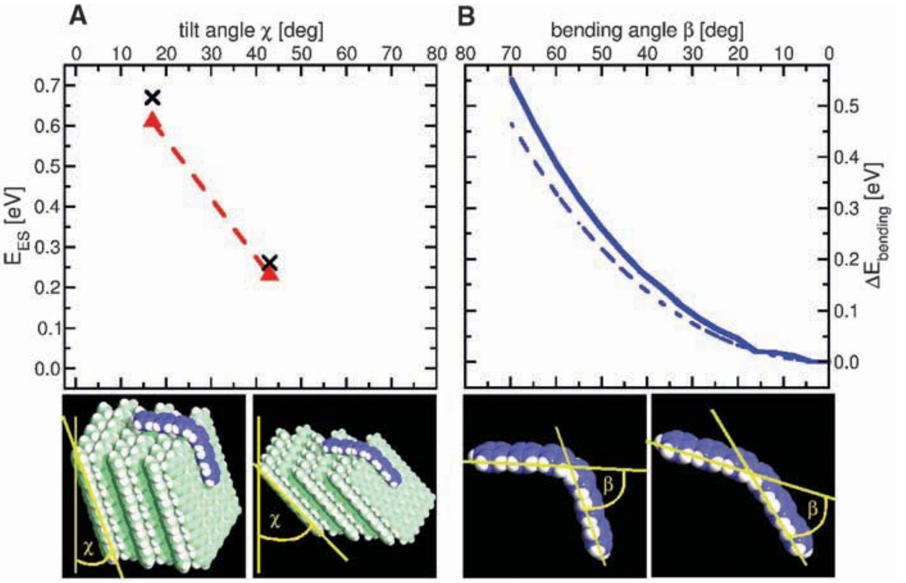


Fig. 4. (**A**) Ehrlich-Schwoebel barrier versus tilt angle of the 6P(001) surface. Experimental values are indicated by black crosses, and theoretical values by red triangles. (**B**) Energy required for bending a single 6P molecule as a function of bending angle. The thick blue line is a result of ab initio density functional theory calculations; the dashed line follows from Brenner's bond order potentials (*30*).

A detailed look at the AFM images of the first few monolayers reveals that the terrace height is not in agreement with the expected bulk value. The morphology of a nominally 1-nm-thick 6P film and the corresponding cross section through one of the islands are shown in Fig. 3, A and B. From these data, we deduce an island height of 2.05 nm ± 0.1 nm. Figure 3C shows the height histogram calculated from a 5-μm image of a 4-nm-thick 6P film. The peaks originate from the uncovered part of each terrace. We can calculate the average terrace height in the individual levels using Gaussians to fit the histogram. The height of the first three layers gradually increases toward the bulk value; cross sections through individual mounds (*16*) confirm this result. The change in terrace height can be explained by a varying tilt angle of the molecules in the first few layers. For the examined films, tilt angles (from the surface normal) between 43° (Fig. 3A) and 17° (bulk) are determined. The situation is sketched for the first four layers in Fig. 3D. Similar observations have already been made for copper phthalocyanine on highly ordered pyrolytic graphite (*29*).

Taking into account the larger tilt angles of the molecules in the first layers, we can calculate the transition state energy using the methods described above. The result of 0.23 eV is again in excellent agreement with the value determined from the analysis of the AFM images. Moreover, we can give a natural explanation for the reduced step-edge barrier. The molecular bending mechanism during interlayer mass transport results from the molecule needing to bend less in the case of the 43° tilted (001) surface. Thus, the required energy barrier for bending decreases, which leads to a rigorous lowering of the step-edge barrier. This effect is illustrated in Fig. 4A, where we plot the dependence of the ESB on the tilt angle of the (001) surface—the ESB depends on the growth stage. Figure 4B shows the bending energy as a function of the bending angle for an isolated molecule and illustrates the observed reduction of the ESB.

The sizable barrier height of 0.67 eV indicates that the growth of smooth films in a layer-by-layer manner is a challenging task for organic semiconductors. However, by carefully controlling the molecular orientation by surfactants or similar means, one might be able to reach that goal more easily than expected. The example of the rodlike 6P molecule shows that growth theories that are well known and verified in inorganic growth can also be used for organic growth. However, we have also demonstrated that the complex nature of the molecular building blocks leads to additional effects that are not observed in atomic inorganic growth.

References and Notes

1. C. K. Chiang *et al.*, *Phys. Rev. Lett.* **39**, 1098 (1977).
2. E. Bauer, *Z. Kristallogr.* **110**, 372 (1958).
3. G. Ehrlich, F. Hudda, *J. Chem. Phys.* **44**, 1039 (1966).
4. R. L. Schwoebel, E. J. Shipsey, *J. Appl. Phys.* **37**, 3682 (1966).
5. M. Klaua, *Rost Krist.* **11**, 65 (1975).
6. K. Meinel, M. Klaua, H. Bethge, *J. Cryst. Growth* **89**, 447 (1988).
7. M. P. Seah, *Surf. Sci.* **32**, 703 (1972).
8. I. Elkinani, J. Villain, *Solid State Commun.* **87**, 105 (1993).
9. M. Kalff, P. Smilauer, G. Comsa, T. Michely, *Surf. Sci.* **426**, L447 (1999).
10. J. Krug, P. Politi, T. Michely, *Phys. Rev. B* **61**, 14037 (2000).
11. T. Michely, J. Krug, *Islands, Mounds and Atoms*, vol. 42 of Springer Series in Surface Science (Springer, Berlin, Heidelberg, 2004).
12. F.-J. Meyer zu Heringdorf, M. Reuter, R. Tromp, *Nature* **412**, 517 (2001).
13. L. Kilian, E. Umbach, M. Sokolowski, *Surf. Sci.* **573**, 359 (2004).
14. S. Zorba, Y. Shapir, Y. Gao, *Phys. Rev. B* **74**, 245410 (2006).
15. M. Fendrich, J. Krug, *Phys. Rev. B* **76**, 121302 (2007).
16. Details of the sample preparation, experiments, and calculations are available at *Science* Online.
17. H. Plank *et al.*, *Thin Solid Films* **443**, 108 (2003).
18. F. Balzer, V. Bordo, A. Simonsen, H.-G. Rubahn, *Phys. Rev. B* **67**, 115408 (2003).
19. C. Teichert *et al.*, *Appl. Phys. A* **82**, 665 (2006).
20. P. Frank *et al.*, *Surf. Sci.* **601**, 2152 (2007).
21. R. Resel, *Thin Solid Films* **433**, 1 (2003).
22. Y.-P. Zhao, H.-N. Yang, G.-C. Wang, T.-M. Lu, *Phys. Rev. B* **57**, 1922 (1998).
23. J. Venables, G. Spiller, M. Hanbücken, *Rep. Prog. Phys.* **47**, 399 (1984).
24. P. Politi, *J. Phys. I* **7**, 797 (1997).
25. G. Koller, S. Surnev, M. Ramsey, F. Netzer, *Surf. Sci.* **559**, L187 (2004).
26. S. Schinzer, S. Köhler, G. Reents, *Eur. Phys. J. B* **15**, 161 (2000).
27. B. Yu, M. Scheffler, *Phys. Rev. B* **55**, 13916 (1997).
28. C. Teichert, C. Ammer, M. Klaua, *Phys. Status Solidi A Appl. Res.* **146**, 223 (1994).
29. H. Yamane *et al.*, *J. Appl. Phys.* **99**, 093705 (2006).
30. D. W. Brenner *et al.*, *J. Phys.* **14**, 783 (2002).

31. This project was supported by the Austrian Science Fund (FWF) Projects S9707, S9714, and P19197.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/108/DC1
Materials and Methods
Figs. S1 to S3
References

22 April 2008; accepted 6 June 2008
10.1126/science.1159455

Large and Rapid Melt-Induced Velocity Changes in the Ablation Zone of the Greenland Ice Sheet

R. S. W. van de Wal,* W. Boot, M. R. van den Broeke, C. J. P. P. Smeets,
C. H. Reijmer, J. J. A. Donker, J. Oerlemans

Continuous Global Positioning System observations reveal rapid and large ice velocity fluctuations in the western ablation zone of the Greenland Ice Sheet. Within days, ice velocity reacts to increased meltwater production and increases by a factor of 4. Such a response is much stronger and much faster than previously reported. Over a longer period of 17 years, annual ice velocities have decreased slightly, which suggests that the englacial hydraulic system adjusts constantly to the variable meltwater input, which results in a more or less constant ice flux over the years. The positive-feedback mechanism between melt rate and ice velocity appears to be a seasonal process that may have only a limited effect on the response of the ice sheet to climate warming over the next decades.

In the recent debate about the sensitivity of the Greenland Ice Sheet to climate warming, much attention has been given to increasing ice velocities. Widespread acceleration of outlet glaciers in southeast Greenland has been detected by satellite radar interferometry (*1*, *2*). These analyses focus on the discharge of outlet glaciers and indicate an increase of mass loss over the first few years of this century (*3*, *4*). Recent work on southeastern outlet glaciers indicates a decrease in the velocities to previous rates associated with a re-equilibration of the calving front (*5*); for Jakobshavn Isbrae along the west side, this is still unclear. In addition to the mass loss by calving, roughly 50% of the annual accumulation is lost by surface melting and runoff in the marginal areas of the ice sheet. Repeat-pass airborne laser altimetry suggests a slight thinning of the ice margin (*6*), although—at present—little ground validation of this finding exists. Concern about the dynamical stability of the ice sheet has also been raised by ground-based Global Positioning System (GPS) measurements at Swiss Camp (69.57°N, 49.31°W, 1175 m above sea level), indicating an additional ice displacement by increased summer velocities with a ratio up to 1.3 between summer and winter velocities (*7*). Clearly, we have to distinguish between mass loss via outlet glaciers and mass loss via increased ablation in the ice marginal zone. Both are affected by increased meltwater production, but the mechanics are different (*8*). On the ice sheet, these interferometry data, with 24-day temporal resolution, revealed summer speed-ups (50 to 100%) consistent with earlier observations (*7*) but smaller seasonal speedups where observed for outlet glaciers (<15%).

Here, we present ice velocity measurements from the major ablation area along the western margin of the ice sheet. The data set contains simultaneous measurements of ice velocity and ablation rates, which makes it possible to study the relation between ice velocity and meltwater input on longer (>5 years) and shorter (~1 day) time scales.

Institute for Marine and Atmospheric research Utrecht, Utrecht University, Netherlands.

*To whom correspondence should be addressed. E-mail: r.s.w.vandewal@uu.nl

The longest continuous time series of surface mass balance observations is that of the K-transect (Fig. 1), starting in 1991 (*9*). This record shows a large year-to-year variability, with ablation rates near the margin varying between 2.5 and 5.6 m of ice per year. The ablation rate in the area is significantly ($P < 0.05$, for linear trend) increasing over time (Fig. 2) but with a large interannual variability, whereas the position of the ice edge is unchanged. In addition to the surface mass balance measurements, stake positions have been recorded annually by GPS. These positions have been converted to annual ice velocities and are presented in Fig. 3A. Annual average ice velocities show a distinct spatial pattern, with the lowest velocities at the higher end of the transect, near Site 10, and the highest velocities at SHR, about 15 km from the ice front. More importantly, it can be observed that ice velocities decrease in time at some locations (Site 5, SHR, and Site 8) and are more or less constant for most other locations. The overall picture obtained by averaging all stake measurements at all sites for individual years indicates a small but significant ($r = 0.79$, $P < 0.05$) decrease of 10% in the annual average velocity over 17 years, which can possibly be attributed to a small decrease in the surface slope or ice thickness. Annually averaged velocities are completely decorrelated to the annual mass balance, whereas a correlation might be expected if there is a strong feedback between velocities and melt rate, leading to enhanced flow, surface lowering, and increased melt rates.

Against this background of slow changes over the past few years, we started more detailed position measurements in 2005 by taking advantage of technological developments of GPS equipment and data processing (*10*). The new instruments record hourly position of stakes, which are drilled into the ice. The GPS (single-frequency) units need to be serviced only once in a year and deliver an ice velocity record with a temporal resolution of 1 day or better.

Ice velocities for the 2005–06 melt season at the site SHR are shown in Fig. 3B. As soon as the surface melt starts in early May, ice velocities increase. Unexpectedly, substantial variations occur, which are superimposed on the gradual increase of the velocities over summer. Changes in ice velocity of up to 30% occur over weekly periods (Fig. 3B). This variability is much stronger than reported previously (*7*). Similar changes in velocity are observed along the entire transect (Fig. 3C). There is a coherent pattern over a distance of 60 km, although the largest changes are at SHR. Earlier in the season, the coherency is restricted to sites located close to each other.

At Sites 5, 6, and 9, automatic weather stations—which measure the ablation rate of the surface with a sonic height ranger—are operated. Combining these data with the high temporal velocity measurements reveals a rapid response of ice velocity to increased melt rates (Fig. 3D). Apparently, the increasing amount of meltwater from the surface penetrates to the base and leads to increased sliding velocities by raising the water pressure near the bed. The response is most prominent at SHR, because this site is located in a zone with many moulins. This finding points to a more general mechanism of meltwater penetration through the ice than the previously reported incidental drainage of a superglacial lake (*11*). Hitherto, the theoretical response of thick ice layers to meltwater fluctuations was assumed to be too slow to lead to weekly variations in the velocity of ice sheets, though observations (*7*) indicated a response. However, recent theoretical work (*12*) shows that if the formation of crevasses can be described by linear elastic fracture mechanics, response times are only of the order of weeks. Our measurements indicate that once the increased velocities are reached in summertime, velocity fluctuations may have an even shorter time scale. An explanation could be that crevasses only partly close during periods of decreased surface melting.

For alpine glaciers, it is known that ice velocities are higher in summer than in winter because of enhanced sliding (*13*, *14*). Local ice thickness on these glaciers is usually only a few hundred meters. Our observations demonstrate that in the ablation zone of the Greenland Ice Sheet, where ice thickness is 1000 to 1500 m, ice velocities respond to changes in the meltwater input within a week, which is in line with but more prominent than the responses observed in previous studies (*7*).

In earlier work (*4*, *7*), it has been suggested that the interaction between meltwater production and ice velocity provides a positive feedback, leading to a more rapid and stronger response of the ice sheet to climate warming than hitherto assumed. Our results are not quite in line with this view. We did not observe a correlation between annual ablation rate and annual ice velocities. Ice velocities respond fast to changes in ablation rate on a weekly time scale. However, on a longer time scale, the internal drainage system seems to adjust to the increased meltwater input in such a way that annual velocities remain fairly constant. In our view, the annual velocities in this part of the ice sheet respond slowly to changes in ice thickness and surface slope.

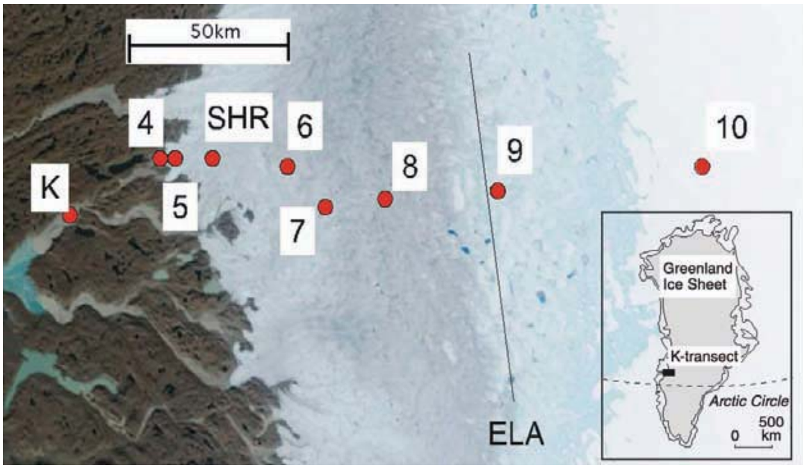
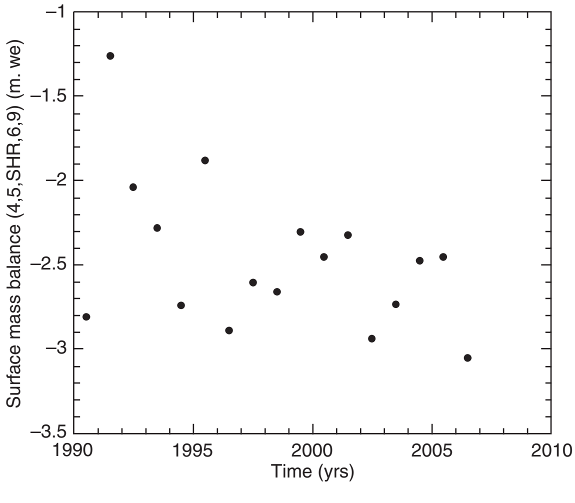


Fig. 1. The K-transect in west Greenland at 67°N. The background NASA–Modis/Terra image is dated 26 August 2003. K is Kangerlussuaq, whereas 4, 5, SHR, 6, 7, 8, and 9 are surface mass balance sites. ELA, Equilibrium Line Altitude. The equilibrium line (indicated by the black line) is at about 1500 m above sea level. The image clearly shows zones, from right to left, of snow (Site 10), wet snow (Site 9), dark ice (Site 8), and clear ice (Sites 4, 5, and SHR).

Fig. 2. The annual surface mass balance in the ablation zone of the K-transect averaged over five stations (Sites 4, 5, SHR, 6, and 9) with a continuous record over the past 17 years. Locations of the individual sites are shown in Fig. 1. "m. we," meter water equivalent.



Longer observational records with high temporal resolution in other ablation areas of the ice sheet are necessary to test the importance of the positive-feedback mechanism between melt rates and ice velocities. At present, we cannot conclude that this feedback is important. We do see a significant increase of the ablation rate (Fig. 2), which is likely related to climate warming, but it remains to be seen if this is likely to be amplified by increasing annual ice velocities.

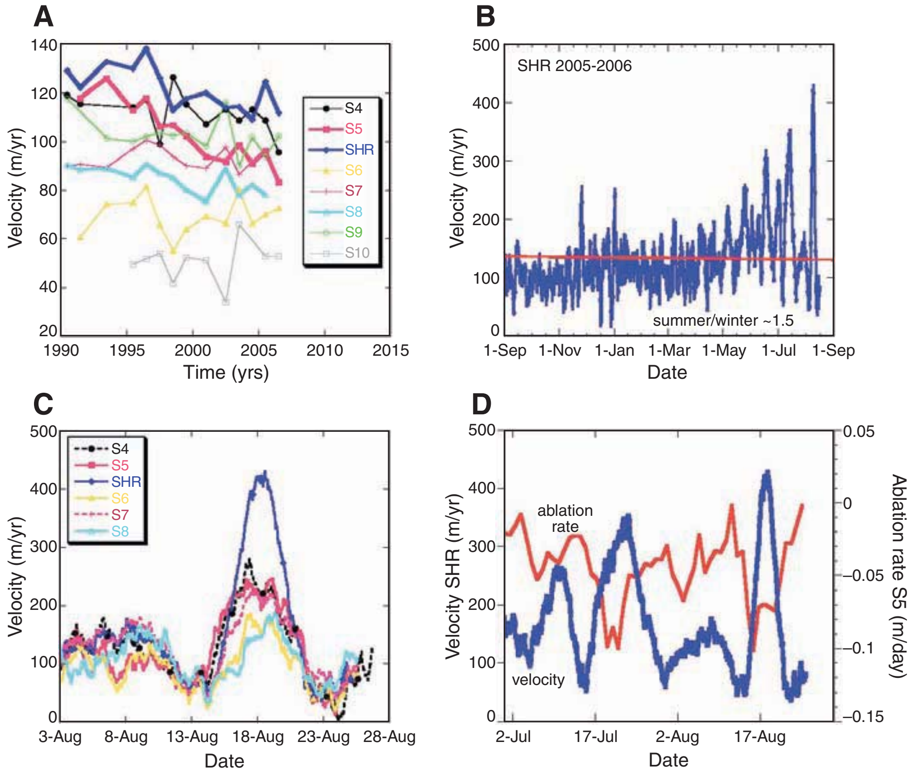


Fig. 3. (**A**) Variations in annual velocity along the K-transect over 17 years; sites with a significant decrease over time are depicted as thick lines. (**B** and **C**) Summer velocities at SHR are about 50% higher than winter velocities (B), which are in phase along the entire transect, particularly at the end of the melt season (C). (**D**) The changes in velocity are clearly related to the ablation rate. If the ablation rate increases, more meltwater is present and velocities increase; however, if ablation ceases, velocities decrease again. This implies that the change in meltwater, rather than the absolute amount of meltwater, determines the change of the velocity within a season.

References and Notes

1. W. Abdalati *et al.*, *J. Geophys. Res.* **106**, 33729 (2001).
2. E. Rignot, P. Kanagaratnam, *Science* **311**, 986 (2006).
3. L. A. Stearns, G. S. Hamilton, *Geophys. Res. Lett.* **34**, L05503 (2007).
4. P. Lemke *et al.*, in *Climate Change 2007*: *The Physical Science Basis. Contribution of Working Group 1 to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007), pp. 337–383.
5. I. M. Howat, I. Joughin, T. A. Scambos, *Science* **315**, 1559 (2007).
6. W. B. Krabill *et al.*, *Geophys. Res. Lett.* **31**, L24402 (2004).
7. H. J. Zwally *et al.*, *Science* **297**, 218 (2002).
8. I. Joughin *et al.*, *Science* **320**, 781 (2008).
9. R. S. W. van de Wal, W. Greuell, M. R. van den Broeke, C. H. Reijmer, J. Oerlemans, *Ann. Glaciol.* **42**, 311 (2005).
10. T. Vincenty, *Surv. Rev.* **23**, 88 (1975).
11. S. B. Das *et al.*, *Science* **320**, 778 (2008).
12. C. J. van der Veen, *Geophys. Res. Lett.* **34**, L01501 (2007).
13. A. Iken, R. A. Bindschadler, *J. Glaciol.* **32**, 101 (1986).
14. T. C. Bartholomaus, R. A. Anderson, S. P. Anderson, *Nat. Geosci.* **1**, 33 (2008).
15. This work was supported by several grants from the Netherlands Organization of Scientific Research and the Netherlands Polar Programme.

1 April 2008; accepted 5 June 2008
10.1126/science.1158540

Mg/Al Ordering in Layered Double Hydroxides Revealed by Multinuclear NMR Spectroscopy

Paul J. Sideris,[1,2] Ulla Gro Nielsen,[1,2]* Zhehong Gan,[3] Clare P. Grey[1,2]†

The anion-exchange ability of layered double hydroxides (LDHs) has been exploited to create materials for use in catalysis, drug delivery, and environmental remediation. The specific cation arrangements in the hydroxide layers of hydrotalcite-like LDHs, of general formula $Mg^{2+}_{1-x}Al^{3+}_{x}OH_2(Anion^{n-}_{x/n})\cdot yH_2O$, have, however, remained elusive, and their elucidation could enhance the functional optimization of these materials. We applied rapid (60 kilohertz) magic angle spinning (MAS) to obtain high-resolution hydrogen-1 nuclear magnetic resonance (^1H NMR) spectra and characterize the magnesium and aluminum distribution. These data, in combination with ^1H-^{27}Al double-resonance and ^{25}Mg triple-quantum MAS NMR data, show that the cations are fully ordered for magnesium:aluminum ratios of 2:1 and that at lower aluminum content, a nonrandom distribution of cations persists, with no Al^{3+}-Al^{3+} close contacts. The application of rapid MAS NMR methods to investigate proton distributions in a wide range of materials is readily envisaged.

Hydrotalcite-like layered double hydroxides (LDHs) are a class of inorganic lamellar compounds with the general chemical composition $M^{2+}_{1-x}M^{3+}_{x}(OH)_2(A^{n-}_{x/n})\cdot yH_2O$, where M^{2+} and M^{3+} are divalent and typically trivalent metal cations respectively, x is the molar ratio of the trivalent cation $[M^{3+}/(M^{2+} + M^{3+})]$, which typically varies between 17% and 33% (*1*), and A^{n-} is an anion with charge n. The presence of a trivalent metal in the metal hydroxide $[M_{1-x}M'_{x}(OH)_2]$ sheet induces an overall positive charge, which is compensated by the incorporation of the anion, along with structural water, in the interlayer spaces (Fig. 1). One naturally occurring example of this class of materials is the mineral hydrotalcite, $Mg_6Al_2(OH)_{16}CO_3{\cdot}4H_2O$, which contains carbonate ions in between the layers. The materials can accommodate a wide range of different anions (*2*) and cations (*3*), leading to a large compositional variety and thus tunability for a large number of applications.

These materials are of considerable geological relevance because of their anion-exchange capacity, which can affect the mobility of chemical species in the environment. Although there is a large group of materials with cation-exchange capabilities, the number of systems with positively charged frameworks or layers is extremely limited. LDHs are, therefore, attractive candidates as anion exchangers and can be used, for example, to remove toxic anions such as chromates (*4*), selenates (*5*), or halides (*6*) from waste waters or, more recently, as drug delivery systems (*7*). The materials are also frequently used as catalysts, catalyst supports, or precursors for oxides used in numerous reactions (1, 8). Organic polymers have been incorporated into LDHs through direct exchange or in situ polymerization from exchanged monomers in the interlayers, creating nanocomposite materials with improved thermal properties (*9*).

An understanding of the basis for the varying anion selectivities of different LDHs requires a molecular-level understanding of the modes of anion binding of the anions within the interlayer spaces. Such insight requires determination of both the structures in the interlayer region and the cation arrangements in the layers themselves, as the latter will control the charge distributions of the hydroxyl groups that point into the interlayer region. Diffraction techniques are not well suited for structural studies of the most prevalent LDHs, namely those containing Mg and Al, for three reasons. First, Mg and Al have effectively identical scattering power, making them indistinguishable by x-ray diffraction. Second, the challenge of synthesizing fully deuterated LDHs, with extremely low residual proton contents, hinders high-resolution neutron studies. Third, many LDHs show considerable turbostratic disorder and stacking faults, complicating the analysis of the diffraction patterns (*10*, *11*).

Nuclear magnetic resonance (NMR) studies of LDHs are particularly attractive due to an abundance of spin-active nuclei from which site-specific structural information can be potentially extracted. However, the high concentration of interlayer water and hydroxyl groups results in strong ^{1}H homonuclear coupling, which broadens the resonances and prevents both ready identification and quantification of chemically distinct ^{1}H environments, particularly at spinning speeds below 25 kHz (*12–14*). ^{27}Al magic angle spinning (MAS) NMR studies of hydrotalcite-like LDHs are the most common and have been used to examine thermal decomposition; however, no clear-cut evidence for cation distributions has emerged from these studies (*15*).

Studies using ^{25}Mg MAS NMR spectroscopy are not routine because of this isotope's low gyromagnetic ratio (which determines sensitivity), low natural abundance (~10%), and large quadrupole moment (which broadens the resonances through the interaction with the local electric-field gradient at the site of the nucleus). Nevertheless, the large quadrupole moment provides a very sensitive probe of the variations in the local environment around Mg, and correlations have, for example, been found between the distortions of the bond angles from perfect octahedral symmetry and the quadrupolar coupling constant C_Q (*16*). Single-pulse ^{25}Mg MAS NMR spectra have been collected for hydrotalcite (*16*, *17*) and the related mineral brucite [$Mg(OH_2)$] (*18–20*), in which tri-octahedral metal hydroxide sheets are occupied solely by Mg. The introduction of Al into brucite-like metal hydroxide sheets produces distortions in the layer, resulting in a significant increase in the electric-field gradient measured at the Mg site. Although the spectra of both minerals could be simulated with a model that assumed only one site geometry [yielding δ_{iso} = 10 parts per million (ppm), C_Q = 4.4 MHz, and δ_{iso} = 13.5 ppm, C_Q = 3.15 MHz, for hydrotalcite and brucite, respectively], it was not clear from this study whether all the different Mg local environments, some of which could be present in low concentrations, were resolved.

Here, we report the results from a combined ^{1}H and ^{25}Mg NMR investigation of three LDHs containing 19%, 25%, and 33% Al [MgAl-19, MgAl-25, and MgAl-33, respectively (*21*)]. Two recent advances in NMR techniques and instrumentation are applied, which noticeably improve the resolution of the spectra acquired from these materials, providing unambiguous evidence for complete cation ordering in the 33% Al doped sample, and a nonrandom distribution of cations in the lower Al-content materials. First, we make use of a MAS NMR probe that is capable of achieving extremely rapid MAS frequencies of up to 60 kHz. At low spinning speeds, broad featureless spectra result because of the large ^{1}H homonuclear dipolar couplings between the OH groups and occluded water molecules (*22*). However, by using MAS frequencies of above 40 kHz, these dipolar couplings are almost completely removed, allowing the individual H sites to be resolved in these high-proton-content materials. Second, we apply two-dimensional (2D) triple-quantum (TQ) ^{25}Mg MAS NMR techniques (*23–25*) to remove the broadening due to the second-order quadrupole interaction. This 2D echo method produces high-resolution spectra—by making use of the different sizes of the second-order quadrupolar interactions in the TQ and single-quantum dimensions of a half-integer spin nucleus such as ^{25}Mg (*26*)—and can be used to reveal multiple sites not evident in the broader 1D spectra. The chosen sample stoichiometries investigated here are representative of low, medium, and high Al content samples, allowing the degree of cation ordering to be explored as a function of Al concentration. Nitrate-containing LDHs are investigated rather than the mineral hydrotalcite, because these materials can then be used for subsequent ion-exchange studies: The carbonate groups in hydrotalcite are not readily exchanged (*27*). However, the approaches demonstrated here can be applied to any Mg- and Al-containing LDHs, and to investigate OH/H_2O environments in a much wider range of minerals.

Once peak broadening is minimized, ^{1}H MAS NMR spectroscopy, in principle, provides a simple method for monitoring cation ordering. Because every hydroxyl group in the LDH layer is coordinated to three metals (each either Mg or Al), four hydroxyl local environments, and thus four distinct ^{1}H resonances, are possible: Mg_3-OH, Mg_2Al-OH, $MgAl_2$-OH, and Al_3-OH (Fig. 1). Assuming a random distribution of the metals on the metal hydroxide sheets, the binomial distribution formula (see SOM text for details) can be used to calculate the percentage of each hydroxyl

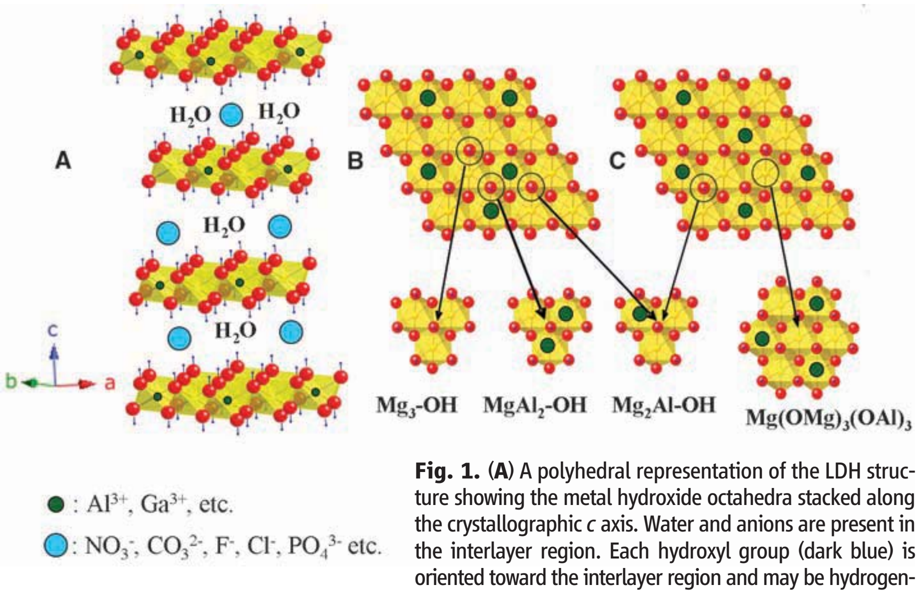

A
H_2O
H_2O
H_2O
H_2O
c
b
a
B
C
Mg_3-OH
$MgAl_2$-OH
Mg_2Al-OH
$Mg(OMg)_3(OAl)_3$
● : Al^{3+}, Ga^{3+}, etc.
● : NO_3^-, CO_3^{2-}, F^-, Cl^-, PO_4^{3-} etc.

Fig. 1. (**A**) A polyhedral representation of the LDH structure showing the metal hydroxide octahedra stacked along the crystallographic *c* axis. Water and anions are present in the interlayer region. Each hydroxyl group (dark blue) is oriented toward the interlayer region and may be hydrogen-bonded to the interlayer anions and water. The metal hydroxide sheets of an LDH with a Mg:Al ratio of 2:1 are shown with (**B**) random and (**C**) ordered cation distributions. Three major classes of hydroxyl groups are present in (B) (Mg_3-OH, Mg_2Al-OH, and $MgAl_2$-OH), whereas only one hydroxyl environment (Mg_2Al-OH) and one Mg local environment [$Mg(OMg)_3(OAl)_3$] are present in (C).

[1] Department of Chemistry, Stony Brook University, Stony Brook, NY 11794–3400, USA. [2] Center for Environmental Molecular Sciences, Stony Brook University, Stony Brook, N 11794–3400, USA. [3] Center of Interdisciplinary Magnetic Resonance, National High Magnetic Field Laboratory, Tallahassee, FL 32310, USA.

*Present address: Department of Physics and Chemistry, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark.

†To whom correspondence should be addressed. E-mail: cgrey@notes.cc.sunysb.edu

resonance as a function of Al content (Table 1). Figure 2A shows the dramatic effect of spinning speed on the resolution of the ^{1}H MAS NMR spectrum of MgAl-33. Three distinct ^{1}H resonances are clearly resolved, which may be readily quantified, when spinning speeds above 40 kHz are used. The relative intensities of these resonances, and their chemical shifts, vary as a function of Al content (Fig. 2B), but the number of distinct resonances remains constant. This observation stands in stark contrast to the random distribution model, which predicts that a total of four hydroxyl groups, and at least one water resonance, will be present.

The spectrum of MgAl-19 shows three ^{1}H resonances at 0.8, 2.4, and 4.7 ppm (Fig. 2B). The chemical shift of the 4.7-ppm resonance is consistent with water occluded in the interlayers (*22*). At low Al concentrations, the two hydroxide environments Mg_3-OH and Mg_2Al-OH should predominate. Given the ~0 ppm chemical shift of the dominant ^{1}H resonance in brucite, which contains only one proton local environment (Mg_3-OH), the resonance at 0.8 ppm is similarly assigned to this local environment. The presence of an Al atom increases the acidity of the hydroxyl group, which results in a shift to higher frequencies (i.e., downfield) (*28*), and on this basis, the peak at 2.4 ppm is assigned to the Mg_2Al-OH group. This resonance shifts gradually to higher frequency as the Al content is increased, which is consistent with stronger hydrogen bonding to the interlayer anions and water with increasing charge on the hydroxide layers. The shift of the Mg_2Al-OH resonance is accompanied by a decrease in the intensity of the 0.8 ppm (Mg_3-OH) resonance, consistent with its assignment. Two-dimensional magnetization exchange experiments (*29*) were performed at a MAS frequency of 60 kHz to explore whether any motion is present in these systems. The experiments reveal that, although slow exchange processes involving exchange between the different OH and H_2O molecules do occur (see SOM text), which are more pronounced for the lower Al-content samples, these do not occur on a time scale that is fast enough to affect the chemical shift positions and intensities of the ^{1}H resonances. Some residual ^{1}H line broadening of the resonances may result from these processes, particularly in the lower Al-content sample MgAl-19.

A large discrepancy exists between the measured relative intensities of the hydroxyl group resonances in the LDH spectra and those calculated assuming the random distribution model, especially for the compound with the greatest Al content, MgAl-33 (Table 1). In particular, no local environments containing more than one Al (e.g., $MgAl_2$-OH and Al_3-OH) are observed, which provides clear evidence that there are no Al^{3+}-Al^{3+} contacts in the hydroxide layers. These trends are consistent with the above assignments of the hydroxyl group resonances and, for MgAl-33, indicate that the Al and Mg cations must be ordered in the metal hydroxide sheets, as shown in Fig. 1C. Furthermore, as shown in Table 1, the experimental concentrations of the hydroxyl groups are in good agreement with the relative concentrations calculated using models that assumed Al-Al avoidance and thus the presence of only Mg_3-OH and Mg_2Al-OH hydroxyl groups. Differences between the experimental intensities and those calculated assuming Al-Al avoidance are ascribed to errors associated with the deconvolution of the ^{1}H NMR spectra and the accurate measurement of Al content by inductively coupled plasma methods (see SOM text for details).

To verify the spectral assignments, we acquired ^{1}H-^{27}Al transfer of population in double-resonance (TRAPDOR) NMR spectra (*30*) of all three samples (Fig. 3). This experiment uses the dipole interaction between ^{1}H and ^{27}Al, which scales as the cube of the through-space distance between the two nuclei. Thus, the most pronounced TRAPDOR effect is expected for the hydroxyl groups coordinated to Al. The TRAPDOR difference spectra of MgAl-19 and MgAl-25 are dominated by a broad resonance whose center of gravity is at ~2.3 ppm, consistent with its assignment to a Mg_2Al-OH group in the ^{1}H MAS spectra. A resonance at 3.5 ppm dominates the difference spectra of MgAl-33, and this shift to higher frequency of the Mg_2Al-OH resonance is again in agreement with the ^{1}H MAS data. The slight discrepancies (≤0.5 ppm) between the centers of gravity of the ^{1}H resonances in the single pulse and ^{1}H-^{27}Al TRAPDOR experiments is ascribed to the much slower spinning frequency used in the TRAPDOR experiments (5 kHz). This results in incomplete averaging of the homonuclear ^{1}H dipolar couplings and thus much poorer resolution of the individual resonances and a loss of signal intensity during the echo experiment, particularly for the strongly coupled ^{1}H spins. Slow motion may also result in some partial loss of signal intensity. The water resonance in these slow spinning spectra is now contained in the intense spinning sidebands spanning over 100 ppm. The most important conclusion from these data are that there are no additional resonances at

Table 1. A comparison of the relative hydroxyl intensities calculated for a random distribution of metals and for an ordered arrangement of cations, assuming Al-Al avoidance, with those determined experimentally from the deconvolution and integration of single-pulse ^{1}H MAS NMR spectra obtained at a spinning speed of 60 kHz. Resonances in the 0.8 to 1.5, and 2.4 to 3.6 ppm ranges contribute to the Mg_3-OH and Mg_2Al-OH local environments, respectively.

	Relative concentrations of hydroxyl groups (%)							
	Random cation distribution				Ordered model		Experimental data	
	Mg_3OH	Mg_2AlOH	$MgAl_2$OH	Al_3OH	Mg_3OH	Mg_2AlOH	Mg_3OH	Mg_2AlOH
MgAl-19	54	37	8	0.6	43	57	38(3)	62(3)
MgAl-25	42	42	14	2	25	75	20(3)	80(3)
MgAl-33	30	44	22	4	1	99	3(1)	97(1)

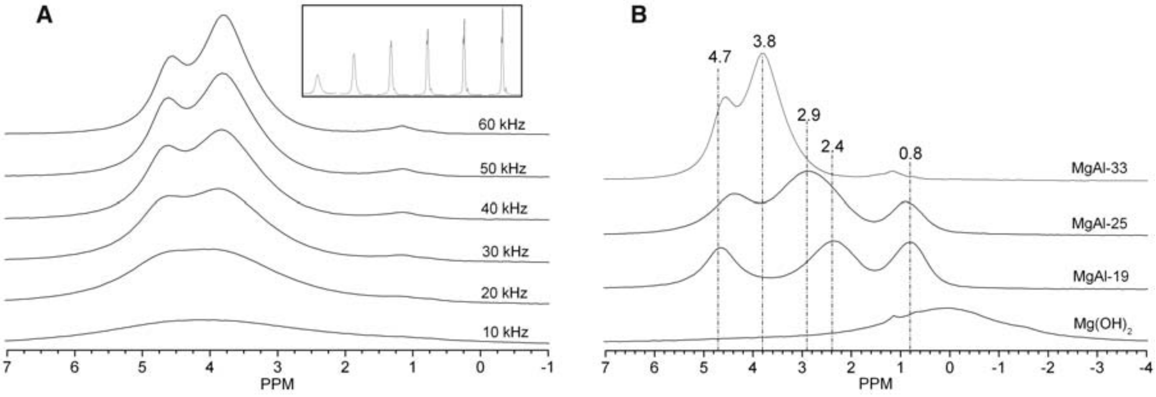


Fig. 2. (**A**) The ^{1}H MAS NMR spectra of MgAl-33 at 14 T as a function of spinning speed: Only the isotropic resonances are shown. The inset shows the same spectra side by side to emphasize the gain in resolution. (**B**) The effect of Al content on the single-pulse ^{1}H MAS NMR spectra collected at a spinning speed of 60 kHz. Hydrotalcite-like LDHs containing 19%, 25%, and 33% Al are shown. The very weak peak [less than 1% of the total signal for $Mg(OH)_2$] at ~1.2 ppm is due to ^{1}H background in the rotor from trace organic impurities (*22*).

higher frequencies, possibly obscured by the water resonance, which can be assigned to environments such as $MgAl_2$-OH.

The ^{25}Mg MAS NMR spectra of brucite, MgAl-19, MgAl-25, and MgAl-33 are shown in Fig. 4A. The spectrum of brucite is dominated by a resonance due to a single, axially symmetric $Mg(OMg)_6$ environment. This brucite-like $[Mg(OMg)_6]$ resonance is present in the spectra of MgAl-19 and MgAl-25, but an additional component, associated with a larger quadrupolar coupling constant, is also observed, which grows in intensity with Al content. The $Mg(OMg)_6$ resonance is absent in the spectrum of MgAl-33, and only the second component is visible, which is again consistent with an axially symmetric Mg local environment.

We acquired TQ MAS NMR spectra to separate the different possible overlapping resonances. The ^{25}Mg TQ MAS NMR spectrum of MgAl-33 (Fig. 4B) clearly shows only one resonance, providing strong evidence that there is only one Mg local environment in this material and that no weaker resonances are hidden under the broad one-pulse spectrum. A projection of the anisotropic dimension was well simulated with the following NMR parameters: $C_Q = 4.6$ (±0.1) MHz, $\delta_{iso} = 13$ (±2) ppm, $\eta = 0.00$ to 0.05; the slight line-shape distortion seen in the experimental spectrum is typical and reflects the orientation dependence of the TQ transition excitation efficiency. The measured C_Q is noticeably different from that of brucite, confirming that the ^{25}Mg electric-field gradient is extremely sensitive to Al substitution. The asymmetry parameter, η, is zero within the error limits, indicating that the Mg ions are in an axial environment, i.e., that there is a C_3 (or higher) rotational axis through the Mg atom. Furthermore, the presence of sharp discontinuities in the line shape implies that a negligible distribution of bond angles and bond lengths, associated with different local environments, is present, as such disorder would asymmetrically broaden the signal.

Only two local environments are consistent with C_3 symmetry, the brucite $Mg(OMg)_6$ environment and the $Mg(OMg)_3(OAl)_3$ environment found in the "honeycomb" ordering of Mg and Al (Fig. 1C). This ordering scheme results in only one local proton environment: Mg_2AlOH. The weak Mg_3OH resonance seen in the ^1H MAS NMR spectrum (accounting for 3% of the hydroxyl groups) of this sample suggests that some Mg-rich defects [e.g., $Mg(OMg)_6$ formed by replacing an Al^{3+} ion by a Mg^{2+} ion in the ordered, honeycomb sheets] are present in the sheets and that the Al content is slightly lower than 33%; these Mg environments are clearly present in too low a concentration to be observed in the TQ MAS spectrum. The ^{25}Mg TQ MAS spectra of the other two LDH samples contains more than one environment for Mg, consistent with their lower Al contents.

In conclusion, a combination of the ^1H MAS data and the ^{25}Mg TQ MAS data clearly shows that the Mg and Al cations are not randomly distributed in the LDH sheets and that in MgAl-33 they are ordered in a honeycomb arrangement. The ordering of the cations affects the charge density of the metal hydroxide sheets, which should have consequences regarding the bonding, reactivity, orientation, and mobility of the chemical species in the interlayer and on the sur-

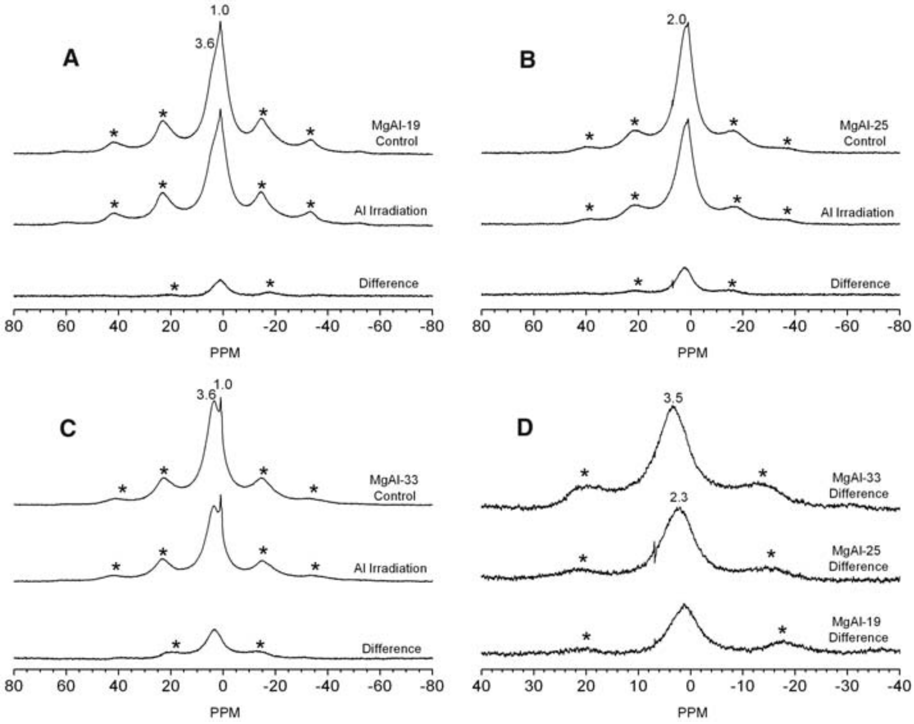


Fig. 3. Results of ^1H-^{27}Al TRAPDOR NMR experiments performed at 8.45 T at a spinning speed of 5 kHz for (**A**) MgAl-19, (**B**) MgAl-25, and (**C**) MgAl-33. The evolution and refocusing period of the spin-echo echo was set to one rotor period (0.2 ms) for all the measurements. The control (no Al irradiation, S_0), double-resonance (Al irradiation, S), and difference spectra (S_0–S), are shown from top to bottom in (A) to (C). A normalized comparison of the difference spectra for all three LDHs is shown in (**D**). * indicates spinning sidebands.

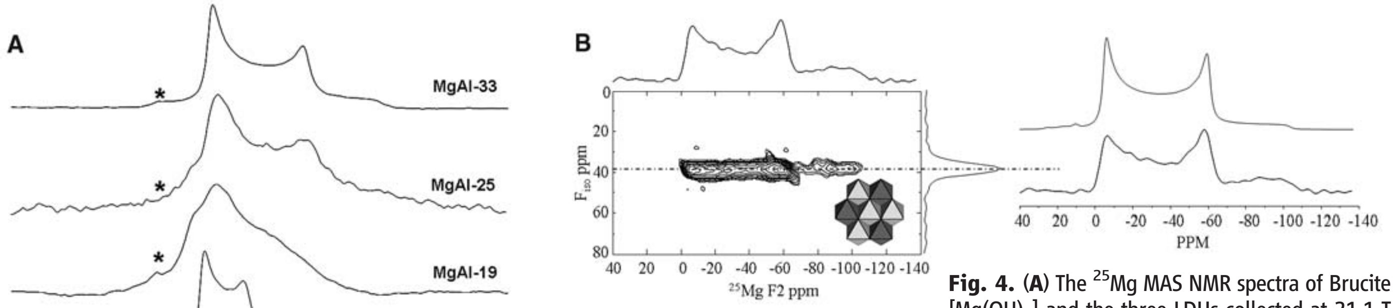


Fig. 4. (**A**) The ^{25}Mg MAS NMR spectra of Brucite $[Mg(OH)_2]$ and the three LDHs collected at 21.1 T at Pacific Northwest National Laboratory. The asterisks indicate the sharper inner satellite transitions, which have a considerably reduced second-order quadrupolar broadening (*31*). (**B**) The ^{25}Mg triple-quantum MAS spectra of ~70% ^{25}Mg-enriched MgAl-33 collected at 19.5 T at the National High Magnetic Field Laboratory. The anisotropic slice, along with a simulation, is shown for the single Mg environment, $Mg(OAl)_3(OMg)_3$.

face. The results clearly demonstrate the power of rapid-spinning ^{1}H MAS NMR spectroscopy to resolve and quantify chemically distinct proton environments in hydrous minerals. This technology provides a simple brute-force method for studying strongly coupled multispin systems that should be straightforward to apply to a wide range of minerals. The approach can readily be used in conjunction with well-established 2D NMR or double-resonance experiments to determine, for example, the relationship between cation ordering and the binding affinities and dynamics of guest anions on the surfaces and in the interlayer spaces.

References and Notes

1. F. Cavani, F. Trifiro, A. Vaccari, *Catal. Today* **11**, 173 (1991).
2. V. Rives, M. A. Ulibarri, *Coord. Chem. Rev.* **181**, 61 (1999).
3. A. I. Khan, D. O'Hare, *J. Mater. Chem.* **12**, 3191 (2002).
4. D. Carriazo, M. del Arco, C. Martin, V. Rives, *Appl. Clay Sci.* **37**, 231 (2007).
5. X. Q. Hou, R. J. Kirkpatrick, *Chem. Mater.* **12**, 1890 (2000).
6. R. P. Bontchev, S. Liu, J. L. Krumhansl, J. Voigt, T. M. Nenoff, *Chem. Mater.* **15**, 3669 (2003).
7. U. Costantino, V. Ambrogi, M. Nocchetti, L. Perioli, *Micropor. Mesopor. Mat.* **107**, 149 (2008).
8. A. Vaccari, *Appl. Clay Sci.* **14**, 161 (1999).
9. F. Leroux, J. P. Besse, *Chem. Mater.* **13**, 3507 (2001).
10. W. Hofmeister, H. Von Platen, *Crys. Rev.* **3**, 3 (1992).
11. A. V. Radha, P. V. Kamath, C. Shivakumara, *Acta Crystallogr. B* **63**, 243 (2007).
12. M. A. Aramendia *et al.*, *J. Solid State Chem.* **131**, 78 (1997).
13. F. Rey, V. Fornes, J. M. Rojo, *J. Chem. Soc., Faraday Trans.* **88**, 2233 (1992).
14. A. Vanderpol, B. L. Mojet, E. Vandeven, E. Deboer, *J. Phys. Chem.* **98**, 4050 (1994).
15. J. Rocha, M. del Arco, V. Rives, M. A. Ulibarri, *J. Mater. Chem.* **9**, 2499 (1999).
16. K. J. D. Mackenzie, R. H. Meinhold, *Am. Mineral.* **79**, 250 (1994).
17. K. J. D. Mackenzie, R. H. Meinhold, B. L. Sherriff, Z. Xu, *J. Mater. Chem.* **3**, 1263 (1993).
18. T. J. Bastow, *Solid State Commun.* **77**, 547 (1991).
19. R. Dupree, M. E. Smith, *J. Chem. Soc. Chem. Commun.* **1988**, 1483 (1988).
20. K. J. D. Mackenzie, R. H. Meinhold, *Thermochim. Acta* **230**, 339 (1993).
21. Materials and methods are available as supporting material on *Science* Online.
22. J. P. Yesinowski, H. Eckert, G. R. Rossman, *J. Am. Chem. Soc.* **110**, 1367 (1988).
23. S. Sham, G. Wu, *Inorg. Chem.* **39**, 4 (2000).
24. C. V. Grant, V. Frydman, L. Frydman, *J. Am. Chem. Soc.* **122**, 11743 (2000).
25. N. G. Dowell, S. E. Ashbrook, S. Wimperis, *J. Phys. Chem. B* **108**, 13292 (2004).
26. L. Frydman, J. S. Harwood, *J. Am. Chem. Soc.* **117**, 5367 (1995).
27. J. He *et al.*, *Struct. Bond.* **119**, 89 (2006).
28. U. Sternberg, E. Brunner, *J. Magn. Res. Ser. A* **108**, 142 (1994).
29. J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *J. Chem. Phys.* **71**, 4546 (1979).
30. C. P. Grey, A. J. Vega, *J. Am. Chem. Soc.* **117**, 8232 (1995).
31. A. Samoson, *Chem. Phys. Lett.* **119**, 29 (1985).
32. Support was provided by the Center for Environmental Molecular Sciences, funded by NSF grant CHE-0021934. Single-pulse ^{25}Mg MAS NMR was performed in the Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at Pacific Northwest National Laboratory. The NMR facility at the National High Magnetic Field Laboratory is supported by NSF and the state of Florida (DMR-0084173). P.J.S thanks J. Ford and A. Palmer for helping with data collection and A. Reynolds for the synthesis of related materials. P.J.S acknowledges the Graduate Assistance in Areas of National Need (GAANN) Fellowship for financial assistance. U.G.N. acknowledges the Camille and Henry Dreyfus Postdoctoral Program in Environmental Chemistry for financial support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/113/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References

10 March 2008; accepted 12 May 2008
10.1126/science.1157581

Autophagy Is Essential for Preimplantation Development of Mouse Embryos

Satoshi Tsukamoto,[1]* Akiko Kuma,[1,2]† Mirei Murakami,[1] Chieko Kishi,[1]
Akitsugu Yamamoto,[3] Noboru Mizushima[1,2]‡

After fertilization, maternal proteins in oocytes are degraded and new proteins encoded by the zygotic genome are synthesized. We found that autophagy, a process for the degradation of cytoplasmic constituents in the lysosome, plays a critical role during this period. Autophagy was triggered by fertilization and up-regulated in early mouse embryos. Autophagy-defective oocytes derived from oocyte-specific Atg5 (autophagy-related 5) knockout mice failed to develop beyond the four- and eight-cell stages if they were fertilized by Atg5-null sperm, but could develop if they were fertilized by wild-type sperm. Protein synthesis rates were reduced in the autophagy-null embryos. Thus, autophagic degradation within early embryos is essential for preimplantation development in mammals.

During the transition from oocyte to embryo, maternal proteins and RNAs are rapidly degraded and replaced by zygotic proteins and RNAs. Degradation of maternal RNAs is thought to be mediated by binding of regulatory proteins to the 3′ untranslated regions of target RNAs or by microRNAs (*1*, *2*). In contrast, the mechanisms by which maternal proteins are degraded remain poorly understood. In mammals, protein degradation accelerates shortly after fertilization and is apparent by the early two-cell stage (*3*). Early embryogenesis may rely on the maternal protein stores as nutrients. Several maternal proteins are degraded by the ubiquitin-proteasome system (*4*, *5*), but it is unknown whether macroautophagy (referred to as autophagy hereafter), another major degradation system, plays an important role during this period. During autophagy, a portion of cytoplasm is sequestered into an autophagosome; this then fuses with the lysosome, and the cytoplasm-derived materials are degraded (*6*–*10*). Autophagy is important for various physiological processes such as starvation adaptation and intracellular quality control.

To investigate whether autophagy occurs in fertilized oocytes, we used autophagy-indicator mice, in which green fluorescent protein (GFP)–fused LC3, a mammalian Atg8 homolog present on autophagosomes, is systemically expressed (*11*, *12*). We collected oocytes and embryos from superovulated GFP-LC3 female mice after mating with wild-type males (*13*). Although metaphase II (MII) oocytes showed almost no GFP-LC3 dots, a number of dots that represented autophagosomes appeared in fertilized embryos at the one- to four-cell stage (Fig. 1, A and B, and fig. S1A). Electron microscopy of two-cell embryos confirmed the presence of autophagic vacuoles (Fig. 1C and fig. S2). Induction of autophagy was also confirmed by LC3 conversion (*11*), which was increased in two-cell embryos relative to MII oocytes (Fig. 1D). Additionally, phosphorylation of S6 kinase was reduced after fertilization, which suggested that mTOR, a negative regulator of autophagy, was inactivated (fig. S3). Thus, formation of auto-

[1] Department of Physiology and Cell Biology, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan. [2] SORST, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan. [3] Department of Bio-Science, Nagahama Institute of Bio-Science and Technology, Nagahama 526-0829, Japan.

*Present address: Laboratory Animal Sciences Section, National Institute of Radiological Sciences, Chiba 263-8555, Japan.

†Present address: Genome Research Institute, University of Cincinnati, Cincinnati, OH 45237, USA.

‡To whom correspondence should be addressed. E-mail: nmizu.phy2@tmd.ac.jp

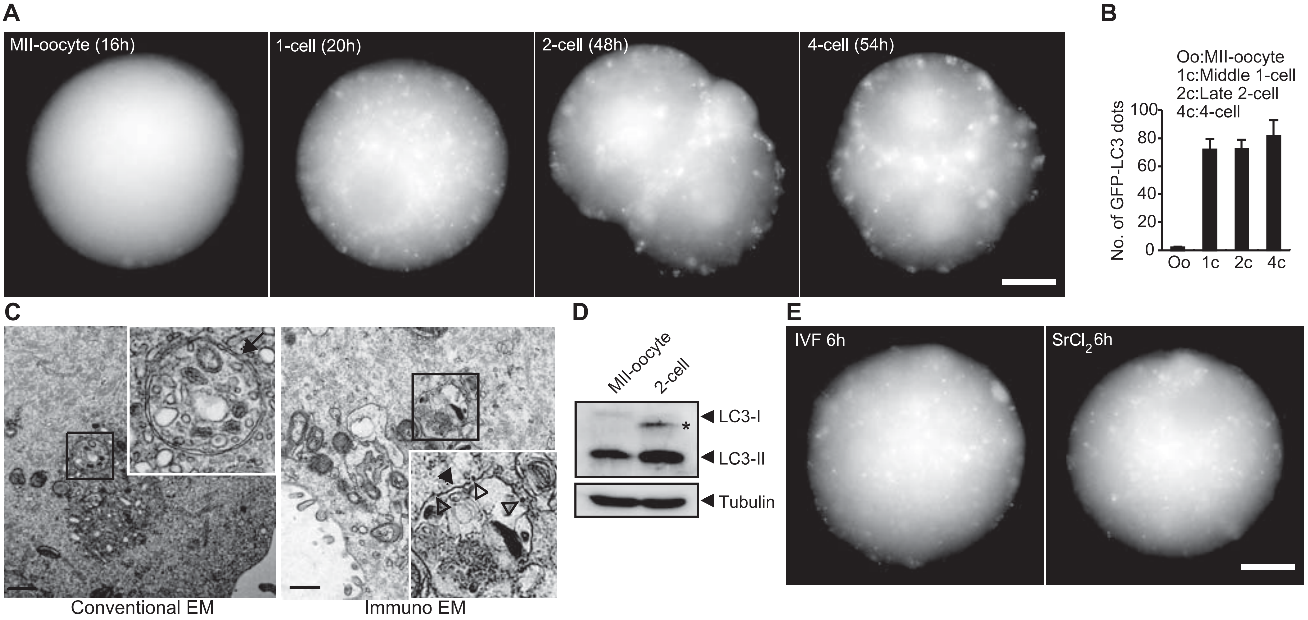
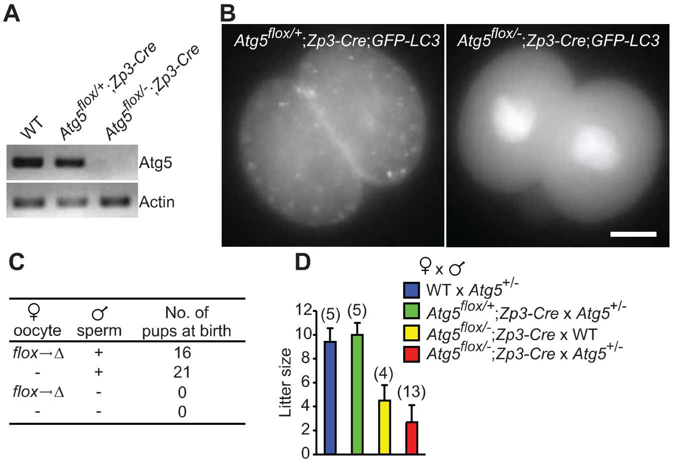


Fig. 1. Induction of autophagy after fertilization. (**A**) Autophagosome formation in preimplantation embryos. MII oocytes and embryos were obtained from superovulated GFP-LC3 transgenic females. Times (in hours) after human chorionic gonadotropin (hCG) treatment are indicated. Scale bar, 20 μm. (**B**) Quantification of GFP-LC3 dots in oocytes and embryos. Numbers of GFP-LC3 dots per embryo image are shown. Each value represents the mean ± SE of at least 10 oocytes or embryos from the indicated stages. (**C**) Conventional and immuno–electron microscopy analysis of embryos at the two-cell stage. An autophagosome and autolysosome are indicated by an arrow and closed arrowhead, respectively. Localization of endogenous LC3 is indicated by open arrowheads. Black squares indicate the enlarged areas shown in insets. Scale bar, 500 nm. (**D**) Conversion of LC3-I (cytosolic) to LC3-II (autophagosome-bound) in fertilized embryos. Whole lysates from 300 MII oocytes and two-cell embryos were loaded and analyzed by immunoblotting. Asterisk indicates an LC3-I degradation product. (**E**) Autophagy induction in embryos 6 hours after IVF or parthenogenetically activated with SrCl$_2$. Scale bar, 20 μm.

phagosomes is up-regulated immediately after fertilization.

Early embryos contained a large number of lysosomes stained with LysoTracker (Molecular Probes) (fig. S1B) (*14*). Some of the GFP-LC3 dots colocalized with LysoTracker, suggesting that autophagosomes had indeed fused with lysosomes. Total expression levels of GFP-LC3 rapidly decreased after the four-cell stage (fig. S1C). Because LC3 in the inner autophagosomal membrane is degraded upon fusion with lysosomes (fig. S1D) (*15*, *16*), the rapid disappearance of GFP-LC3 protein suggests that autophagic flux is up-regulated in fertilized embryos.



♀ oocyte	♂ sperm	No. of pups at birth
flox→Δ	+	16
-	+	21
flox→Δ	-	0
-	-	0

Fig. 2. Post-fertilization autophagy is essential for development. (**A**) Generation of oocyte-specific Atg5-deficient mice, as shown by reverse transcription polymerase chain reaction analysis of Atg5 and actin mRNA levels in MII oocytes derived from wild-type, *Atg5$^{flox/+}$;Zp3-Cre*, and *Atg5$^{flox/–}$;Zp3-Cre* female mice. (**B**) Absence of GFP-LC3 dot formation in Atg5-deficient embryos. Embryos from *Atg5$^{flox/+}$;Zp3-Cre;GFP-LC3* and *Atg5$^{flox/–}$;Zp3-Cre;GFP-LC3* females crossed with wild-type males were isolated at 48 hours after hCG injection. Scale bar, 20 μm. (**C**) Genotypic distribution of offspring obtained from *Atg5$^{flox/–}$;Zp3-Cre* females crossed with *Atg5$^{+/–}$* males. (**D**) Average litter size of indicated mating pairs. Numbers above each bar are numbers of pregnancies. Each value represents the mean ± SD.

We next determined whether autophagy is induced simply as a result of starvation after ovulation. MII oocytes were collected from GFP-LC3 mice and inseminated in vitro with wild-type sperm. The number of GFP-LC3 dots remained very small until 3 hours, but numerous dots appeared by 4 hours after in vitro fertilization (IVF) and increased in number thereafter (Fig. 1E and movie S1). Autophagy was not induced in unfertilized oocytes cultured under the same conditions. We also observed that unfertilized oocytes in the oviduct showed only very small numbers of GFP-LC3 dots, which suggests that the occurrence of autophagy depends on fertilization rather than ovulation (fig. S1F). After induction, autophagy was transiently suppressed between the late one-cell stage and the middle of the two-cell stage, and was then reactivated (fig. S1E). Transient suppression of autophagy has also been observed during mitosis in various mammalian cell lines (*17*).

We then sought to determine whether fertilization itself or subsequent events are sufficient to induce autophagy. Oocytes isolated from GFP-LC3 mice were parthenogenetically activated with strontium and cultured in vitro. In these oocytes, GFP-LC3 dot generation was observed with a time course similar to that observed for in vitro fertilized embryos (Fig. 1E). Because strontium induces repetitive intracellular calcium release in a manner similar to that seen after normal fertilization (*18*), calcium oscillation may be an inducer of autophagy.

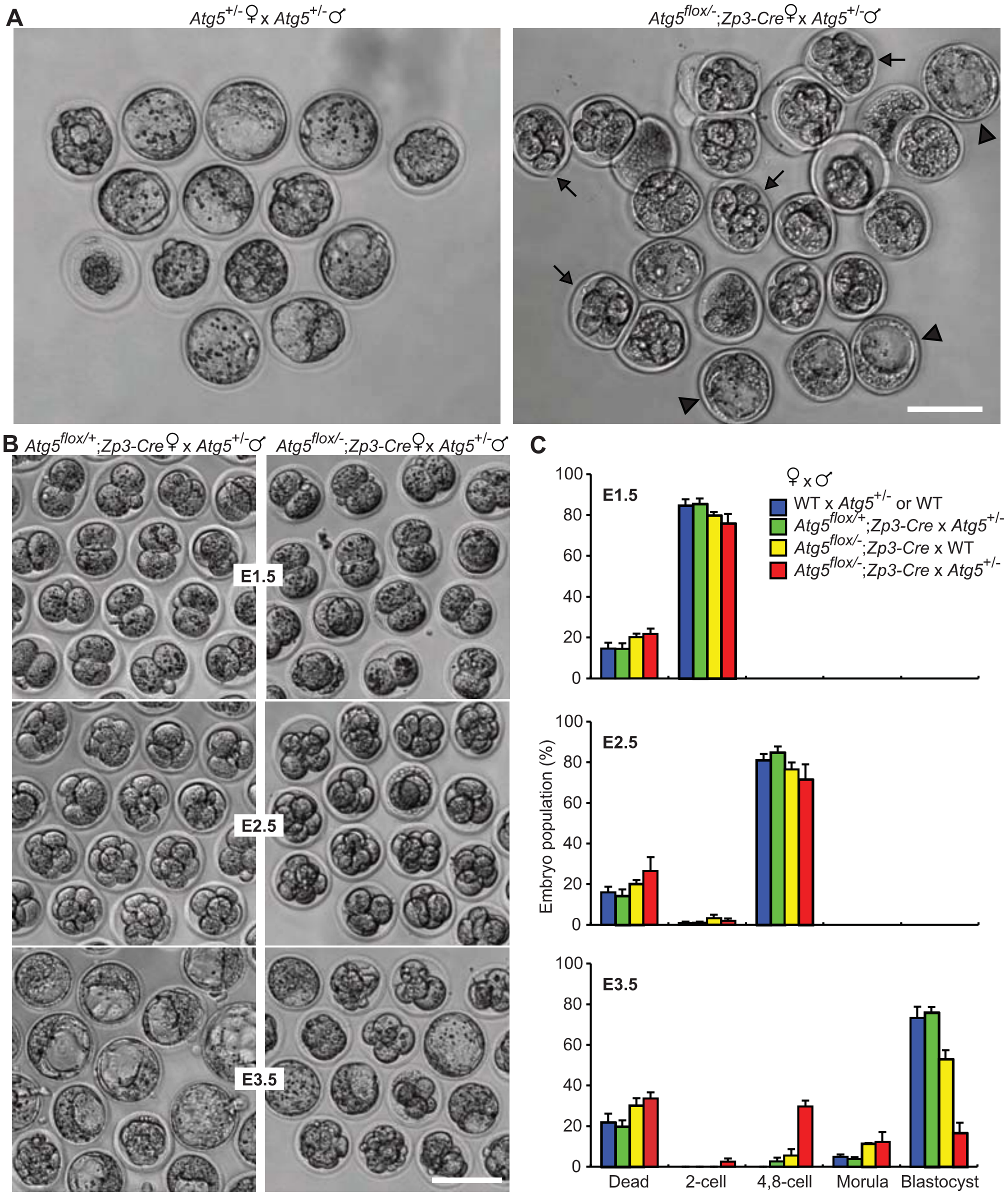


Fig. 3. Autophagy deficiency causes embryonic lethality at the four- to eight-cell stage. (**A**) Defective in vivo development of autophagy-deficient embryos. E3.5 embryos from *Atg5*$^{+/-}$ or *Atg5*$^{flox/-}$*;Zp3-Cre* females mated with *Atg5*$^{+/-}$ males are shown. Arrowheads indicate normally developing blastocysts; arrows indicate developmentally retarded embryos. Scale bar, 100 μm. (**B**) Defective in vitro development of autophagy-deficient embryos. Embryos were collected from *Atg5*$^{flox/+}$*;Zp3-Cre* and *Atg5*$^{flox/-}$*;Zp3-Cre* females mated with *Atg5*$^{+/-}$ or wild-type males at the two-cell stage (E1.5) and subsequently cultured in vitro for 2 days until the blastocyst stage (E3.5). Representative bright-field photographs of embryos at each stage are shown. Scale bar, 100 μm. (**C**) Distributions of embryonic developmental stage at each time point. Data represent means ± SE of three to five different experiments.

Atg5 is an essential factor for autophagosome formation (*19*). *Atg5*$^{-/-}$ mice appear grossly normal at birth but die within 1 day of delivery, which implies that autophagy may not be important during embryogenesis (*20*). However, these *Atg5*$^{-/-}$ mice were generated by mating *Atg5*$^{+/-}$ mice, so maternally inherited Atg5 protein in the cytoplasm of Atg5-null oocytes might have rescued the autophagy-deficient phenotype during early embryogenesis. Indeed, we observed normal levels of autophagy in *Atg5*$^{-/-}$ two-cell embryos derived from *Atg5*$^{+/-}$ mice (fig. S4). This is consistent with our previous observation that only very small amounts of Atg5 are required for autophagy (*21*). To analyze the functional relevance of autophagy during early embryogenesis, we generated oocyte-specific Atg5-deficient mice, in which both the *Atg5* gene and Atg5 protein could be deleted in oocytes. Mice bearing an *Atg5*^{flox} allele (*22*) were crossed to a transgenic line expressing Cre recombinase under the control of the Zp3 promoter, which is active in growing oocytes (Zp3-Cre mice) (*23*).

The ovaries of *Atg5*$^{flox/-}$*;Zp3-Cre* mice contained all developmental stages of oocyte and were morphologically indistinguishable from ovaries observed in *Atg5*$^{flox/+}$*;Zp3-Cre* mice (fig. S5A). Normal numbers of oocytes were recovered from superovulated *Atg5*$^{flox/-}$*;Zp3-Cre* females (fig. S5B) and could be fertilized normally (fig. S5C). *Atg5* mRNA expression was completely suppressed in MII oocytes collected from *Atg5*$^{flox/-}$*;Zp3-Cre* females (Fig. 2A). Thus, autophagy is not essential for oogenesis or fertilization. We next determined whether autophagy was induced in these oocytes after fertilization. In two-cell embryos collected from *Atg5*$^{flox/-}$*;Zp3-Cre;GFP-LC3* females that had been crossed with wild-type males, GFP-LC3 dots were not observed; in contrast, GFP-LC3 dots were extensively generated in embryos from *Atg5*$^{flox/+}$*;Zp3-Cre* (Fig. 2B) and wild-type females (Fig. 1A). In the autophagy-deficient two-cell embryos, GFP-LC3 accumulated in the nucleus; similar events are often observed in other cell types under autophagy-inactive conditions, although the physiological importance is unknown (*24*). Thus, oocytes from *Atg5*$^{flox/-}$*;Zp3-Cre* females are completely defective in autophagy, and this cannot be rescued at the two-cell stage by wild-type sperm.

We then examined in vivo development of the oocyte-specific Atg5-deficient mice. After mating of *Atg5*$^{flox/-}$*;Zp3-Cre* females and *Atg5*$^{+/-}$ males, we obtained only *Atg5*$^{\Delta/+}$ and *Atg5*$^{-/+}$ (Δ indicates the floxed allele successfully rearranged by the Cre recombinase) pups; we did not obtain either *Atg5*$^{\Delta/-}$ or *Atg5*$^{-/-}$ (Fig. 2C and fig. S5D) pups. Thus, fertilization of Atg5-null oocytes with Atg5-null sperm resulted in embryonic lethality.

In contrast, Atg5-null oocytes were able to produce pups when fertilized with wild-type sperm, which suggests that the deficiency in autophagy can be rescued by zygote-derived Atg5. We thus determined the stage at which rescue started. In embryos derived from *Atg5*$^{flox/-}$*;Zp3-Cre;GFP-LC3* females mated with wild-type males, zygotic *Atg5* mRNA was detected as early as the two-cell stage (fig. S6A). However, induction of autophagy was not observed immediately; large numbers of GFP-LC3 dots appeared only after the eight-cell stage (fig. S6B). At the eight-cell stage, the nuclear accumulation of GFP-LC3 also disappeared. The timing of rescue seemed late, given the rapid onset of post-fertilization autophagy in wild-type embryos (Fig. 1A). Consistently, we noticed that the litter sizes of *Atg5*$^{flox/-}$*;Zp3-Cre* females mated with *Atg5*$^{+/+}$ males (Fig. 2D, yellow) were smaller than those of *Atg5*$^{flox/+}$*;Zp3-Cre* or wild-type females crossed with *Atg5*$^{+/-}$ males (Fig. 2D, blue and green). Thus, restoration of autophagy at the eight-cell stage is insufficient to completely rescue embryonic viability. However, we cannot exclude the possibility that, despite normal ovulation and fertilization rates, autophagy-deficient oocytes are slightly compromised, resulting in low birth rate. In either case, the clear dependence of survival on sperm genotype (Fig. 2C) suggests a critical role of autophagy after, rather than before, fertilization.

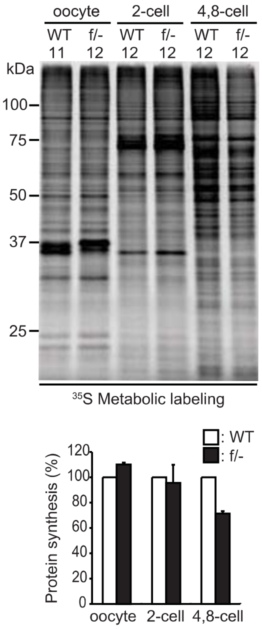


Fig. 4. Reduced protein synthesis in autophagy-defective embryos. Oocytes and embryos were radiolabeled with [^{35}S]methionine for 2 hours. Total lysates were resolved by SDS–polyacrylamide gel electrophoresis and subjected to autoradiography. ^{35}S incorporation was measured and normalized against signals from the wild type. Numbers of oocytes or embryos used are shown above each lane. Data represent means ± SE of three independent experiments. WT, embryos derived from wild-type females crossed with wild-type males; f/-, embryos derived from *Atg5*$^{flox/-}$*;Zp3-Cre* females crossed with *Atg5*$^{+/-}$ males.

We next determined the stage at which development was perturbed in Atg5-deficient embryos by both in vivo and in vitro experiments. *Atg5*$^{flox/-}$*;Zp3-Cre* females were mated with *Atg5*$^{+/-}$ males, and embryos were collected and examined at various stages. Development seemed to proceed normally until the four-cell stage; however, when embryos were collected at embryonic day 3.5 (E3.5), fewer than half of the embryos reached the blastocyst stage, while most of the control embryos became blastocysts (Fig. 3A). Even after culturing these E3.5 embryos for an additional 24 hours in vitro, the abnormal embryos did not develop and remained at the four- or eight-cell stage (fig. S7).

For the in vitro experiments, E1.5 embryos were collected and cultured for 2 days. Again, embryos obtained from *Atg5*$^{flox/-}$*;Zp3-Cre* females crossed with *Atg5*$^{+/-}$ males developed to the four- to eight-cell stage almost normally (Fig. 3B; Fig. 3C, red). However, after 48 hours in culture, only 17% of embryos reached the blastocyst stage (all of which were *Atg5*$^{\Delta/+}$ or *Atg5*$^{+/-}$, fig. S8), and 30% of embryos (some of which appeared fragmented) remained at the four- to eight-cell stage (Fig. 3, B and C, E3.5). Reduced numbers of blastocysts (53%) were detected at 48 hours in matings of *Atg5*$^{flox/-}$*;Zp3-Cre* females and wild-type males (Fig. 3C, yellow), consistent with the results above (Fig. 2D). Taken together, these data suggest that autophagy deficiency causes a developmental block at the four- to eight-cell stage.

One possible mechanism underlying the developmental defects of autophagy-deficient embryos might be impaired protein recycling. Although there was no apparent difference in protein synthesis rates up to the two-cell stage, [^{35}S]methionine incorporation into proteins in autophagy-defective embryos was reduced to about 70% of that of wild-type embryos at the four- and eight-cell stages (Fig. 4). Suppression of protein synthesis to similar levels with cycloheximide also caused a developmental block, confirming the importance of new protein synthesis during this period (fig. S9). The impaired protein synthesis in autophagy-deficient embryos is likely due to amino acid insufficiency; however, we cannot exclude the possibility that the protein synthesis defects are a secondary effect of a role of autophagy in the removal of obsolete maternal factors, or for energy production within embryos.

Our results show that autophagy is upregulated shortly after fertilization and is essential for preimplantation development. These results are not inconsistent with the finding that conventional *Atg5*$^{-/-}$ mice survive preimplantation development (*20*). Rather, these two different mouse models clearly demonstrate the specific importance of autophagy during very early post-fertilization development, in which maternally inherited Atg5 protein remains in the cytoplasm. Autophagy may be dispensable for later development. We do not know whether post-fertilization autophagy is specifically important for mammals. In contrast to birds, fish, and amphibians, mammalian preimplantation development progresses very slowly without extracellular nutrient stores; thus, autophagy may be a unique strategy to support mammalian development.

References and Notes

1. W. Tadros, H. D. Lipshitz, *Dev. Dyn.* **232**, 593 (2005).
2. A. F. Schier, *Science* **316**, 406 (2007).
3. E. A. Merz, R. L. Brinster, S. Brunner, H. Y. Chen, *J. Reprod. Fertil.* **61**, 415 (1981).
4. C. DeRenzo, G. Seydoux, *Trends Cell Biol.* **14**, 420 (2004).
5. M. L. Stitzel, G. Seydoux, *Science* **316**, 407 (2007).
6. B. Levine, D. J. Klionsky, *Dev. Cell* **6**, 463 (2004).
7. A. M. Cuervo, *Trends Cell Biol.* **14**, 70 (2004).
8. D. J. Klionsky, *Nat. Rev. Mol. Cell Biol.* **8**, 931 (2007).
9. N. Mizushima, *Genes Dev.* **21**, 2861 (2007).
10. N. Mizushima, B. Levine, A. M. Cuervo, D. J. Klionsky, *Nature* **451**, 1069 (2008).
11. Y. Kabeya *et al.*, *EMBO J.* **19**, 5720 (2000).
12. N. Mizushima, A. Yamamoto, M. Matsui, T. Yoshimori, Y. Ohsumi, *Mol. Biol. Cell* **15**, 1101 (2004).
13. See supporting material on *Science* Online.
14. G. Sun-Wada *et al.*, *Dev. Biol.* **228**, 315 (2000).
15. I. Tanida, N. Minematsu-Ikeguchi, T. Ueno, E. Kominami, *Autophagy* **1**, 84 (2005).
16. E. Shvets, E. Fass, Z. Elazar, *Autophagy* **4**, 621 (2008).
17. E.-L. Eskelinen *et al.*, *Traffic* **3**, 878 (2002).
18. A. Bos-Mikich, K. Swann, D. G. Whittingham, *Mol. Reprod. Dev.* **41**, 84 (1995).
19. N. Mizushima *et al.*, *J. Cell Biol.* **152**, 657 (2001).
20. A. Kuma *et al.*, *Nature* **432**, 1032 (2004).
21. N. Hosokawa, Y. Hara, N. Mizushima, *FEBS Lett.* **580**, 2623 (2006).
22. T. Hara *et al.*, *Nature* **441**, 885 (2006).
23. W. N. de Vries *et al.*, *Genesis* **26**, 110 (2000).
24. A. Kuma, M. Matsui, N. Mizushima, *Autophagy* **3**, 323 (2007).
25. We thank M. Fukuda for antibody to LC3, and N. Minami and F. Aoki for helpful discussions and critical reading of the manuscript. Supported in part by Grants-in-aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan. We also thank the Kato Memorial Bioscience Foundation and the Toray Science Foundation for financial support.

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/117/DC1
Materials and Methods
Figs. S1 to S9
Movie S1
References

4 January 2008; accepted 3 June 2008
10.1126/science.1154822

Phylogenetic Signal in the Eukaryotic Tree of Life

Michael J. Sanderson

Molecular sequence data have been sampled from 10% of all species known to science. Although it is not yet feasible to assemble these data into a single phylogenetic tree of life, it is possible to quantify how much phylogenetic signal is present. Analysis of 14,289 phylogenies built from 2.6 million sequences in GenBank suggests that signal is strong in vertebrates and specific groups of nonvertebrate model organisms. Across eukaryotes, however, although phylogenetic evidence is very broadly distributed, for the average species in the database it is equivalent to less than one well-supported gene tree. This analysis shows that a stronger sampling effort aimed at genomic depth, in addition to taxonomic breadth, will be required to build high-resolution phylogenetic trees at this scale.

Reconstruction of the phylogenetic history of a large sample of life on Earth is nearly within reach. Molecular sequence data are available in GenBank for 10% of described species diversity (*1*); improvements in algorithms and high-performance computing technology have dramatically increased the scale of feasible phylogenetic inference (*2*); and unconventional sources of data, including whole genomes (*3*), expressed sequence tag libraries (*4*, *5*), and barcode sequences (*6*), have altered the landscape of large-scale phylogenetics with an infusion of new evidence. The pace of phylogenetic discovery has accelerated to the point where nearly complete phylogenetic trees can be constructed for well-studied clades, such as mammals (*7*). Such high-resolution trees—those including all taxa for which data are available—permit strong inferences regarding problems ranging from conservation biology (*8*) to comparative biology (*9*) to reconstructing ancestral genomes (*10*). The phylogenetic distribution of species in GenBank is remarkably broad, as a visualization of the National Center for Biotechnology Information (NCBI) taxonomy tree (*11*, *12*) shows (Fig. 1 and figs. S1 and S2). Construction of a high-resolution phylogenetic tree containing all eukaryotic species in the database is a grand challenge that is substantially more tractable than inferring the entire tree of life, but to succeed, strategies will have to overcome serious sampling impediments. Quantifying the distribution and strength of phylogenetic evidence currently in the database is a prerequisite for this effort.

The NCBI taxonomy tree provides a convenient framework for organizing a series of phylogenetic analyses of the 181,992 eukaryotic taxa having sequence data, henceforth termed operational taxonomic units (OTUs) (*13*). I partitioned the NCBI tree into 1127 higher taxa at the rank of order for further analysis {table S1 [see (*13*) for an alternative rank-free partition, which yielded similar results]}. I then downloaded from release 1.01 of the PhyLoTA Browser database [(*14*, *15*) on the basis of GenBank release 159] 14,289 potentially phylogenetically informative clusters of homologous sequences assembled for each higher taxon. Each cluster has a minimum of four OTUs, which is necessary to provide resolution in an unrooted tree. Unrooted phylogenetic trees were constructed for each cluster with a fast but conservative (*16*, *17*) procedure taking both alignment and phylogenetic uncertainty into account. Any clade in the resulting tree will have had at least 50% bootstrap support in maximum parsimony "fast" bootstrap analyses (*18*) with two different sequence alignment algorithms (*19*, *20*). Although this protocol biases the confidence assessment slightly downward, the bias is small (*13*). Of greater concern is that the sequence data used here are enriched for taxonomic diversity to the relative exclusion of some high-throughput genomics data (*13*), which, though presently available for only a small fraction of eukaryotic taxa, ultimately should enable stronger phylogenetic inferences.

Phylogenetic support for each cluster was measured by the fraction of clades resolved on the final alignment-merged bootstrap consensus tree [its consensus fork index (*13*)]. Phylogenetic support for each OTU in the NCBI tree was measured by the sum of the support measures of all phylogenetically informative sequence clusters that contained it. Finally, phylogenetic support for a higher taxon, *H*, was measured by the mean support score of the OTUs contained within it, which is the weighted sum, $\bar{s}_H = \frac{\sum_k n_k(H) w_k(H)}{n_H}$, where $w_k(H)$ is the support score for cluster *k* in higher taxon *H*, $n_k(H)$ is the number of OTUs in cluster *k*, and n_H is the number of OTUs in *H*. This support score was selected among many possibilities in part because of its relative insensitivity to the size of the higher taxon [Fig. 2; see supporting online material (SOM) text]. For comparative purposes and to aid in the visualization of results, an arbitrary cutoff value of 1.5 was selected as

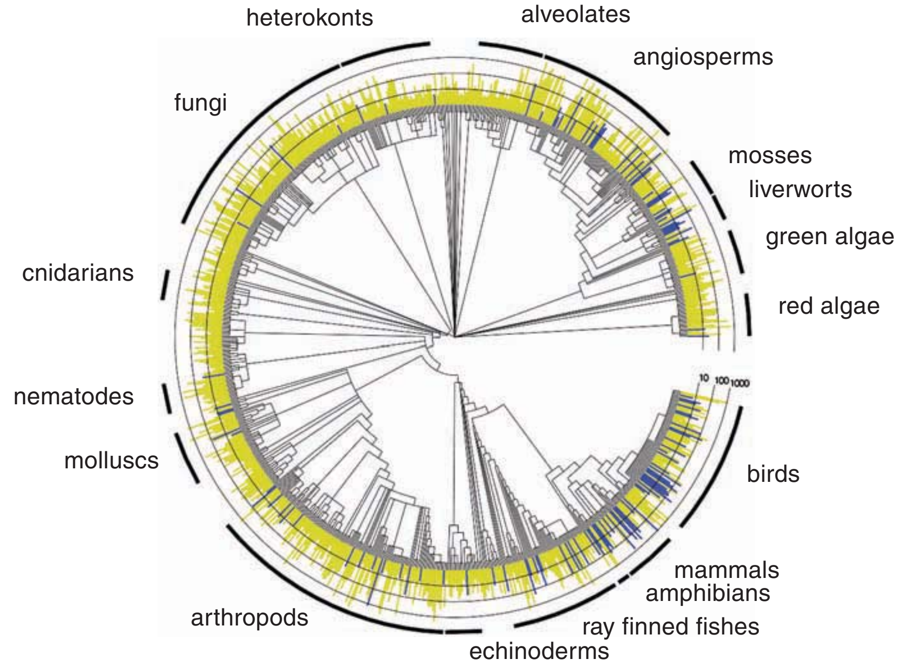


Fig. 1. Phylogenetic support across the NCBI taxonomy tree of eukaryotes. The tree displays 876 taxonomic orders. Not shown are 251 orders from the original selection of 1127 with fewer than four OTUs, which could not contain phylogenetically informative clusters by definition (*13*). Rectangles for each order are colored blue if they exhibit minimum phylogenetic support (1.5 support units or higher) or yellow if they do not. The radial length of these rectangles is proportional to the log of diversity (number of OTUs) within that order (black circles provide scale for diversity). Arcs are labeled with selected major eukaryotic clades. See fig. S1 for a high-resolution image of this figure showing all ordinal names. See the SOM for parallel results for a rank-free partition of the NCBI tree.

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. E-mail: sanderm@email.arizona.edu

minimal phylogenetic support. This is equivalent, for example, to the information content of two independent loci, each resolving three-quarters of clades to at least a bootstrap value of 51%.

Among individual OTUs, *Homo sapiens* had the maximum support value of 293.9, but the distribution of scores had a long tail leading to 6402 OTUs with no support at all (most of which, 6079, simply were not found in any phylogenetically informative clusters). The top 10 were all mammals; the top 25 were mammals, angiosperms (tomato, potato, tobacco, rice, and wheat), *Drosophila melanogaster*, and *Drosophila simulans*, all with support scores above 60 units. Of the 171,703 OTUs for which scores were calculated, only 12% achieved minimal phylogenetic support. The mean support was 0.84, less than the equivalent of each taxon being found in at least one well-resolved and -supported phylogenetic tree.

The mean support values of orders were skewed (tables S4 and S5), ranging from a maximum of 10.0 in primates to 0.0 in 75 other orders (mean = 0.93 among the 876 with at least four OTUs; mean number of clusters = 1.88). The phylogenetic position of orders with minimal support is shown in Fig. 1, as well as the order's species richness. A very similar picture emerges in the rank-free partition of the NCBI tree (fig. S2). In the ordinal partition, only 14% of higher taxa achieved minimal support. The support within those orders in the NCBI tree that happened to be species-rich (≥100 OTUs) provides a good indication of potential for high-resolution tree inference (tables S2 and S3). Their phylogenetic position appears nonrandom. In vertebrates, 16 of 57 species-rich orders had minimal support, but across arthropods only 1 out of 45 did, the acalyptrates (containing *Drosophila*), and none of the other species-rich metazoan orders did. Fungi, with 40 diverse orders, had only 1 order that achieved minimal phylogenetic support; and angiosperms, with 45 species-rich orders, had only 3, still far short of vertebrate performance. Many areas of the NCBI tree, including the vast diversity of metazoa that are neither vertebrate nor arthropod, as well as the diversity of microbial eukaryotes, have few orders, species-rich or -poor, that achieved minimal phylogenetic support. Some taxa with surprisingly low support exemplify how biological diversity can overwhelm substantial and sustained phylogenetic efforts. Examples include the legumes and grasses in angiosperms, two groups containing the bulk of the most economically important and well-studied plants on Earth (*21*, *22*); and the huge subgroup of ascomycete fungi, the Pezizomycotina (*23*, *24*), containing numerous plant and animal pathogens, sources of antibiotics such as penicillin, and organisms used in human food production.

The finding that the average eukaryotic species or higher taxon in GenBank has a phylogenetic support score of less than 1.0 units (10 times less than the best-supported vertebrate orders) has several implications. An accurate high-resolution phylogeny will require substantial increases in sequence data to bring that score to a level comparable to that of the best-supported higher taxa. Although improved phylogenetic inference tools, such as new methods of inferring species trees from collections of gene trees (*25*, *26*), may ultimately extract more power from the same quantity of data, new sampling strategies will also be needed to both acquire and warehouse specimens for DNA work (such as in DNA banks at major natural history collections) and to survey the largest number of relevant genomic sequences per sample. Sampling efficiency can be improved dramatically by targeting the addition of new sequences to the right clusters. One target is the large number of currently phylogenetically uninformative clusters that would become informative with the addition of just a few sequences. Loci can also be targeted in newly acquired species by paying attention to the size and support value of clusters already constructed for related taxa, a practice followed informally by systematists but which can now be quantified with some precision (*13*). For example, in the angiosperm clade Solanales, the five clusters contributing most to support are, in decreasing order, nuclear ribosomal DNA internal transcribed spacers, plastid *ndhF*, nuclear *GBSS*, the plastid *trnL* spacer region, and plastid *rbcL*, so these are obvious targets for further sampling. Recent advances in applying information theory may make possible more-nuanced sampling algorithms that take into account cluster sequence variation, number of taxa, and phylogenetic depth of the tree (*27*). None of these considerations address the difficult sampling issue of undescribed species, most of which lie in the regions of Fig. 1 that are already least well-supported. These are not just absent from GenBank and Fig. 1, they are unknown to science. They can only be added through more biodiversity surveys and alpha taxonomic work. In the meantime, sampling protocols guided by quantitative assessments of the phylogenetic distribution of data will improve the efficiency of emerging phylogenomic strategies for building the tree of life of known organisms.

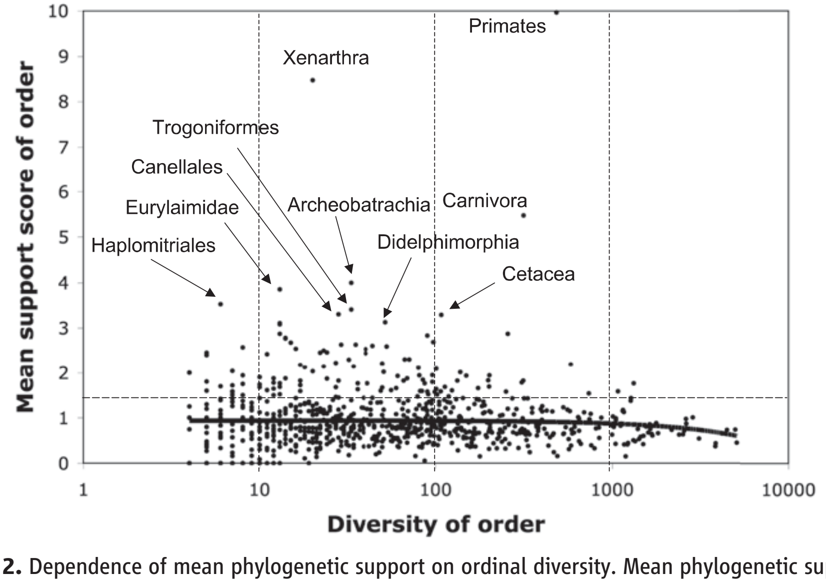


Fig. 2. Dependence of mean phylogenetic support on ordinal diversity. Mean phylogenetic support in taxonomic orders is not dependent on the diversity of these taxa [n = 876, P = 0.18 (not significant), R^2 = 0.002: The solid line is a linear regression, which is slightly curved in the graph's log scale], which suggests that the variation seen across eukaryotes is due not so much to the size of the order but to its phylogenetic position. The horizontal dotted line is at a support value of 1.5 units, the level corresponding to minimum phylogenetic support in this study. Vertical dotted lines are placed at diversities of 10, 100, and 1000 OTUs to correspond with the diversity scale in Fig. 1. Taxon names for the 10 highest-scoring orders are indicated. Two are land plants (Canellales, an angiosperm; and Haplomitriales, a liverwort order); the other eight are all vertebrates (two bird, one amphibian, and five mammalian orders).

References and Notes

1. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler, *Nucleic Acids Res.* **35**, D21 (2007).
2. M. J. Sanderson, *Aust. Syst. Bot.* **20**, 287 (2007).
3. A. G. Clark *et al.*, *Nature* **450**, 203 (2007).
4. J. Hughes *et al.*, *Mol. Biol. Evol.* **23**, 268 (2006).
5. D. Steinke, W. Salzburger, A. Meyer, *J. Mol. Evol.* **62**, 772 (2006).
6. M. Hajibabaei, G. A. C. Singer, P. D. N. Hebert, D. A. Hickey, *Trends Genet.* **23**, 167 (2007).
7. O. R. P. Bininda-Emonds *et al.*, *Nature* **446**, 507 (2007).
8. F. Forest *et al.*, *Nature* **445**, 757 (2007).
9. A. Moles *et al.*, *Science* **307**, 576 (2005).
10. M. Blanchette, E. D. Green, W. Miller, D. Haussler, *Genome Res.* **14**, 2412 (2004).
11. S. Federhen, in *The NCBI Handbook*, 2003 (www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.ch4).
12. NCBI Taxonomy Browser, 2007 (www.ncbi.nlm.nih.gov/Taxonomy/txstat.cgi).
13. Detailed materials and methods are available as supporting material on *Science* Online.
14. M. J. Sanderson, PhyLoTA Browser, 2007 (http://loco.biosci.arizona.edu/cgi-bin/pb.cgi).
15. M. J. Sanderson, D. Boss, D. Chen, K. A. Cranston, A. Wehe, *Syst. Biol.* **57**, 35 (2008).
16. M. E. Mort, P. S. Soltis, D. E. Soltis, M. L. Mabry, *Syst. Biol.* **49**, 160 (2000).

17. N. Salamin, M. W. Chase, T. R. Hodkinson, V. Savolainen, *Mol. Phylogenet. Evol.* **27**, 528 (2003).
18. D. L. Swofford, *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA, ed. 4.0, 2002).
19. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
20. R. C. Edgar, *Nucleic Acids Res.* **32**, 1792 (2004).
21. M. Wojciechowski, M. Lavin, M. Sanderson, *Am. J. Bot.* **91**, 1846 (2004).
22. T. Kajita, H. Ohashi, Y. Tateishi, C. Bailey, J. Doyle, *Syst. Bot.* **26**, 515 (2001).
23. J. W. Spatafora *et al.*, *Mycologia* **98**, 1018 (2006).
24. T. Y. James *et al.*, *Nature* **443**, 818 (2006).
25. C. Ané, B. Larget, D. A. Baum, S. D. Smith, A. Rokas, *Mol. Biol. Evol.* **24**, 412 (2007).
26. S. V. Edwards, L. Liu, D. K. Pearl, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5936 (2007).
27. E. Mossel, M. Steel, in *Mathematics of Evolution and Phylogeny* (Oxford Univ. Press, New York, 2005), pp. 384–412.
28. I thank B. Boyle, K. Cranston, D. Hearn, M. McMahon, B. O'Meara, and T. Wheeler for critical discussions; three anonymous reviewers for critical comments; D. Boss for computing support; and NSF and the University of Arizona BIO5 Institute for financial support.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/121/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S5

20 December 2007; accepted 31 March 2008
10.1126/science.1154449

Accelerated Human Population Growth at Protected Area Edges

George Wittemyer,*† Paul Elsen, William T. Bean, A. Coleman O. Burton, Justin S. Brashares*†

Protected areas (PAs) have long been criticized as creations of and for an elite few, where associated costs, but few benefits, are borne by marginalized rural communities. Contrary to predictions of this argument, we found that average human population growth rates on the borders of 306 PAs in 45 countries in Africa and Latin America were nearly double average rural growth, suggesting that PAs attract, rather than repel, human settlement. Higher population growth on PA edges is evident across ecoregions, countries, and continents and is correlated positively with international donor investment in national conservation programs and an index of park-related funding. These findings provide insight on the value of PAs for local people, but also highlight a looming threat to PA effectiveness and biodiversity conservation.

The past three decades have seen a 500% increase in land designated as protected areas (PAs) for nature conservation (*1*). Many see this explosion of land protection as negatively impacting the livelihoods of local communities through a loss of rights, exclusion from natural resources, and displacement from traditional lands (*2*–*4*). As a result, emigration from protected lands may be expected to reduce human population growth at PA edges over time relative to neighboring unprotected lands, if we assume that higher intrinsic growth rates (often linked to increased poverty levels) are not associated with PA proximity. However, PA creation may benefit rural inhabitants by providing access to road networks, employment, foreign aid, increasingly scarce ecosystem services (e.g., firewood, bushmeat, and clean water), and areas of safety during strife (*5*) (Table 1). If PAs are perceived locally to provide opportunities otherwise scarce in rural areas, we might expect immigration to drive high rates of population growth along PA borders as suggested by a number of recent case studies (*6*, *7*). To investigate this question, we quantified rates of population growth around 306 PAs in 45 countries of Africa and Latin America (Fig. 1).

Using spatially explicit population data for Africa and Latin America from each decade between 1960 and 2000 (*8*, *9*), we calculated average annual rates of population growth within a 10-km buffer of 306 rural (*10*) International Union for Conservation of Nature (IUCN) category I and II PAs and nature World Heritage Sites (*1*) (Fig. 1 and table S1). We then compared the mean rates of growth in buffers to national rural population growth rates as measured by the United Nations in the countries in which the PAs were located (*11*). Our results show average annual growth rates were higher in PA buffers than in rural areas of the same country for 245 of the 306 PAs and in 38 of 45 countries (Figs. 1 and 2). Results were similar across continents, although strongest in Latin America, with buffers around 149 of 164 PAs and in 14 of 16 countries demonstrating greater average growth rates as compared with 96 of 142 PAs and 24 of 29 countries in Africa (Fig. 1). Comparisons of growth rates in PA buffers to alternative estimates of average rural population growth produced similar results (*10*).

Our comparison of population growth around the borders of PAs with average rural rates for the same country (*11*) may present a false picture of human settlement if parks are preferentially placed in areas of high ecological productivity. In such a scenario, humans may settle in the same general region as PAs simply because the land there is better for agriculture or natural resource extraction rather than for reasons related to the PA itself (*12*). To account for this possibility, we refined our analysis by restricting our comparison of population growth rates in the buffers of PAs to those areas with the same ecological characteristics, defined using the Global Ecoregions Database (*13*). Results of this comparison show that, similar to our countrywide comparisons, human population growth around PAs is significantly higher than that observed in matched ecoregions (Wilcoxon test: $Z = -291.5$, $n = 69$, $P = 0.04$).

It also is conceivable that the observed high rates of human population growth in PA buffers are caused by the displacement of people living within PAs to their edges (*3*). In such a scenario, population growth within parks should decline over time as people move outwards toward PA edges. However, contrary to this expectation, population growth rates were positive, not negative, inside 85% of the PAs we surveyed with the remaining 15% showing no change. This finding makes clear that "leakage" from within parks does not explain our result, as population growth was positive not only at PA borders but also within PAs.

A number of social and economic factors may explain accelerated population growth on PA edges (Table 1). We suggest that this pattern is explained by immigration, but if PAs are located in relatively

Table 1. Potential attractants and deterrents of PAs as settlement sites.

Attractants	Deterrents
Foreign aid and integrated conservation and development projects (e.g., schools and clinics)	Land-use restrictions
	Wildlife conflict (e.g., livestock and crop depredation)
Employment (e.g., staff positions and tourism)	Conflict with park staff, government, or rural militias
Enhanced ecosystem services (e.g., wood, food, water, and traditional medicine)	Cultural degradation and increased cost of living associated with tourism
Market access (e.g., road infrastructure)	
Security (e.g., guards and government staff)	Isolation and/or remoteness from urban centers

Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: georgew@nature.berkeley.edu (G.W.); brashares@nature.berkeley.edu (J.S.B.)

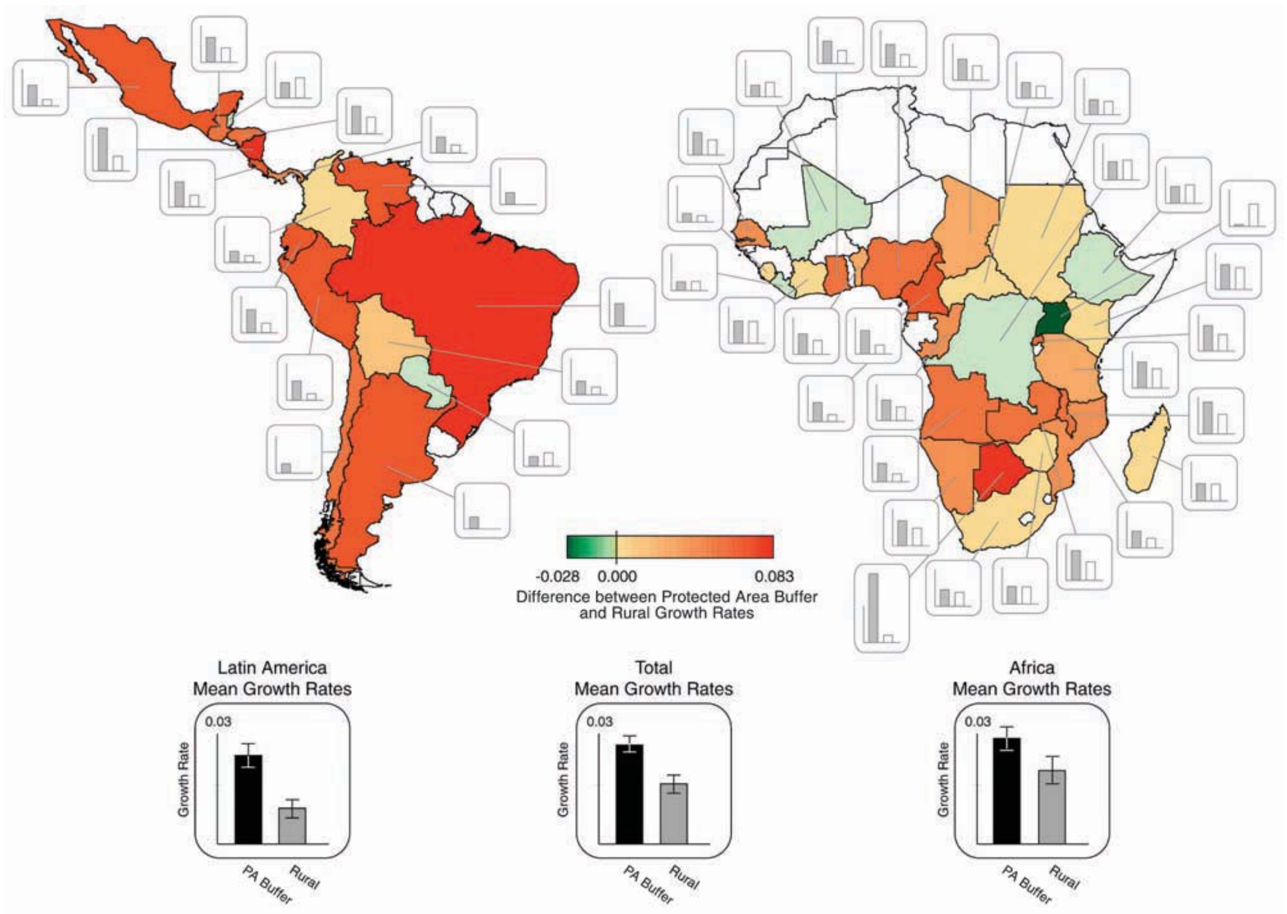

-0.028
0.000
0.083
Difference between Protected Area Buffer and Rural Growth Rates
Latin America Mean Growth Rates
Total Mean Growth Rates
Africa Mean Growth Rates
0.03
Growth Rate
PA Buffer
Rural

Fig. 1. Across Africa and Latin America, human population growth rates in 10-km buffers surrounding PAs (black bars) nearly doubled those of national rural growth rates (gray bars), exceeding them by an average of ~1% per annum. Error bars show standard errors of the means. Buffer growth rates topped the national rural growth rate in approximately 85% of the 45 countries for which PAs were assessed (colors are scaled by the difference between buffer and rural growth rates). The smaller histograms compare average buffer (gray) and rural (white) population growth rates for each country. Growth rates in PA buffers were unrelated to PA (*29*) size (Spearman's rank correlation: r_s = –0.05, n = 284, P = 0.40); country size (r_s = –0.21, n = 45, P = 0.16); or the proportion of area under protection in a country (r_s = 0.23, n = 45, P = 0.12).

impoverished regions of a country or if PAs promote impoverishment, the recognized link between poverty and intrinsic population growth may better explain the patterns we observe. To evaluate this alternative explanation, we examined published data on infant mortality rates (*14*), a widely recognized correlate of poverty, in areas adjacent to and more distant from PAs in 34 of the 45 countries included in our study (data from 11 countries were not available). We found that rates of infant mortality did not differ between PA edges and other rural areas in these countries (Wilcoxon test: Z = –20, n = 34, P = 0.36), which suggests that poverty rates are not higher near PAs. This result and observations that population growth around PAs sometimes exceeded maximum human reproductive rates support our contention that immigration drives the patterns we report.

Few data are available to allow many broad-scale and rigorous tests of the role of political, economic, and ecological drivers of human settlement around PAs. Nevertheless, limited data for a subset of 126 of our focal PAs (*15*) show that the difference between PA buffer growth and rural growth was positively correlated with the density of PA staff (employees per hectare) (Spearman's rank correlation: r_s = 0.19, n = 126, P = 0.03). Because staff levels are shown to be positively linked with PA budgets and tourism rates (*16*), PA staff density may serve as an index of a PA's enforcement activity or its economic solvency. As such, the positive correlation we observe indicates population growth rates in buffers are likely influenced by increased economic or occupational opportunities, rather than suppressed through exclusion from natural resources found within PAs. In practice, the mechanisms driving population changes around PAs are likely context-specific, and data collection at local scales is critical for understanding the relation between local people and PAs.

Our observation of high population growth along PA borders in Africa and Latin America may not be surprising considering the significant park-focused integrated conservation and rural development investment made by international donor agencies (*7*, *17*, *18*). For example, from 1991 to 2006, the Global Environment Facility (GEF) and its funding partners alone distributed on the order of $2 billion to support PAs in Africa and Latin America (*19*). In fact, for countries included in our study, the amount the GEF spent on biodiversity projects (*20*) during this period was positively correlated with the rates at which PA buffer growth exceeded rural growth [Spearman's rank correlation: r_s = 0.38, n = 36, P = 0.02; (fig. S1)]. This correlation could simply show that GEF invests most heavily in countries where population growth around PAs is highest, but we think it more likely that international investment contributes to the settlement patterns we observe. We also found a positive correlation between per capita Gross Domestic Product (GDP) and the rates at which PA buffer growth exceeded rural growth in the 45 developing countries included in our analysis (r_s = 0.38, n = 45, P = 0.01). If GDP is a reliable indicator of a country's investment in PA development, population growth around PAs may reflect a demographic response to both national and international funding.

Full accounts are difficult to obtain, but even by conservative estimates, the 306 PAs included in our study have received millions of dollars from international development and conservation organizations since their creation (*16*). This fund-

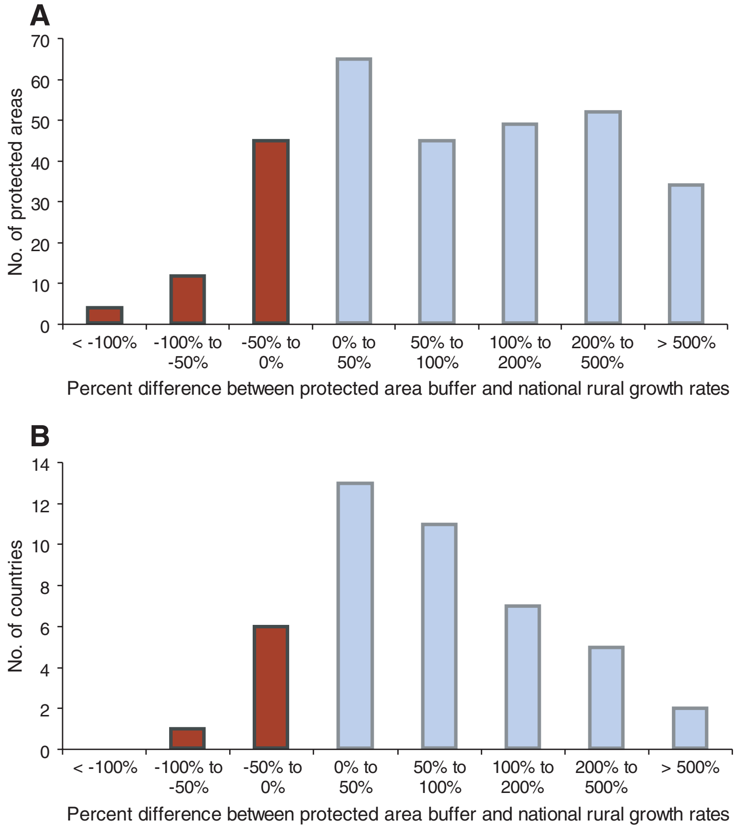


Fig. 2. Population growth rates within 10 km of 306 rural PAs in 45 countries in Africa and Latin America (*29*) were significantly higher than national rural averages (*11*) when tested by (**A**) PA (Wilcoxon test: $Z = -1.58 \times 10^4$, $n = 284$, $P < 0.001$) or (**B**) country ($Z = -270$, $n = 36$, $P < 0.001$).

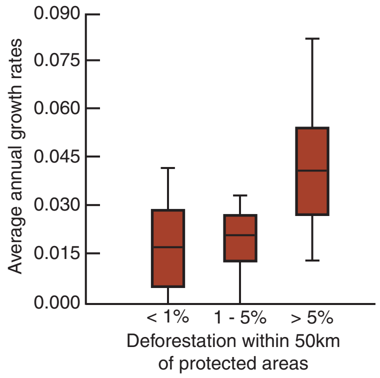


Fig. 3. Deforestation measured within a 50-km buffer of 55 forested PAs in Africa and Latin America (*26*) was greatest where annual rates of human population growth in PA buffers were highest (Kruskal-Wallis test: $H_2 = 16.48$, $P < 0.001$). Box plots show the median (horizontal line), interquartile range (box), and range of the data (whiskers). The two lower deforestation categories showed no difference with regard to population growth rates (rank-sum difference test: $P = 0.27$). Population growth in buffers where deforestation was greater than 5% was significantly higher than in the two lower deforestation categories, both independently and combined (rank-sum difference tests: $P < 0.001$).

ing includes development and capacity-building projects for people living in PA buffers where it directly and indirectly benefits communities through the creation of jobs, roads, clinics, sanitation systems, and schools (*19*, *21*). In some cases, such park-centered financial infusion far exceeds international funds targeted for development in communities distant from PAs. As such, the perceived benefits of park-related development, along with access to increasingly scarce ecosystem services, may be expected to drive immigration and settlement of people on PA edges, a concern that has been a long-standing topic of debate in the development and conservation communities (*6*, *7*). If we assume that patterns of human settlement reflect the will and prosperity of local people, our results suggest local populations in Africa and Latin America perceive benefits from living in proximity to PAs.

Although PAs may be positive for localized rural development in Africa and Latin America, human populations around PAs frequently have significant, negative impacts on biodiversity (*22*). The scale of human settlement around PAs is a strong predictor of illegal timber and mineral extraction (*23*), bushmeat hunting (*24*), fire frequency (*25*), and, more generally, species extinction (*24*) within PAs. We examined such impacts directly by comparing population growth rates in PA buffers with published rates of deforestation in the area surrounding 55 forest PAs included in our study (*26*). Rates of deforestation were highest around PAs where human population growth was greatest (Fig. 3). This finding links population growth around PAs to habitat loss and suggests settlement around PAs may create a ring of disturbance that isolates PAs from surrounding habitats. Although population growth along PA edges clearly has impacted tropical forest habitats, our analyses across all habitat types indicate PAs in grassland and dry forest ecosystems may be the most at risk as buffers in these regions contain particularly high population growth rates (fig. S2). If PAs are expected to serve as refuges for the "last of the wild" (*27*), the patterns we document here are cause for concern.

If humans are drawn to PAs for the economic opportunities they provide, international funding for conservation may, ironically, exacerbate the same anthropogenic threats to biodiversity it aims to alleviate. By no means should this possibility rationalize reduced funding for PAs and the communities around them. Instead, it suggests that international and local funding must go in part toward developing spatially dynamic PA systems that account for patterns of human settlement and needs of local communities. Creation of large multi-use buffer areas surrounding core habitats and corridors (possibly with mixed-use buffers of their own) between PAs may facilitate effective protection of biodiversity while supporting potentially heavy human settlement on PA borders. Additionally, approaches that pair PA-based conservation with economic development targeted at areas more distant from PAs may aid rural communities while simultaneously reducing human pressure on PAs (*28*). Such advanced landscape planning, in concert with effective PA management, may maintain and increase the benefits of PAs for rural people while also ensuring those benefits do not result in unsustainably heavy use of the flora, fauna, and processes PAs endeavor to sustain.

References and Notes

1. U.N. Environment Programme (UNEP)–World Conservation and Monitoring Center (WCMC) and IUCN, *World Database on Protected Areas* [IUCN–World Commission on Protected Areas (WCPA)/UNEP-WCMC, Washington, DC, 2003]; available at www.unep-wcmc.org/wdpa/ (accessed September 2007).
2. W. M. Adams *et al.*, *Science* **306**, 1146 (2004).
3. M. M. Cernea, K. Schmidt-Soltau, *World Dev.* **34**, 1808 (2006).
4. N. L. Peluso, *Glob. Environ. Change* **3**, 199 (1993).
5. L. M. Scherl *et al.*, *Can Protected Areas Contribute to Poverty Reduction? Opportunities and Limitations* (IUCN, Gland, Switzerland, 2004).
6. A. de Sherbinin, M. Freudenberger, *Parks* **8**, 38 (1998).
7. P. Scholte, *Ambio* **32**, 58 (2003).
8. UNEP International Center for Tropical Agriculture and the World Resource Institute, *Latin America Population Distribution Database* [UNEP/Global Resource Information Database (UNEP/GRID)–Sioux Falls, Sioux Falls, SD, 2004]; available at www.na.unep.net/datasets/datalist.php (accessed September 2007).
9. UNEP and Center for International Earth Science Information Network (CIESIN), *Africa Population Distribution Database* (UNEP/GRID–Sioux Falls, Sioux Falls, SD, 2004); available at www.na.unep.net/datasets/datalist.php (accessed September 2007).

10. Materials and methods are available as supporting material on *Science* Online.
11. U.N. Population Division (UNPD), *World Urbanization Prospects: The 2007 Revision Population Database* (UNPD, New York, 2007); available at http://esa.un.org/unup (accessed September 2007).
12. R. P. Cincotta, J. Wisnewski, R. Engelman, *Nature* **404**, 990 (2000).
13. D. M. Olson *et al.*, *Bioscience* **51**, 933 (2001).
14. A. de Sherbinin, *Oryx* **42**, 26 (2008).
15. UNEP-WCMC, *Prototype Nationally Designated Protected Areas Database* (WCMC/UNEP, Cambridge, 2003); available at www.unep-wcmc.org/protected_areas/data/nat.htm (accessed September 2007).
16. A. N. James, M. J. B. Green, J. R. Paine, *Global Review of Protected Area Budgets and Staff* (WCMC, Cambridge, 1999).
17. A. Balmford, K. J. Gaston, S. Blyth, A. James, V. Kapos, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1046 (2003).
18. A. James, K. J. Gaston, A. Balmford, *Bioscience* **51**, 43 (2001).
19. GEF, *GEF Biodiversity Strategy in Action* (GEF, Washington, DC, 2006).
20. GEF, *The GEF Project Database* (GEF, Washington, DC, 2008); available at http://gefonline.org/home.cfm (accessed March 2008).
21. W. D. Newmark, J. L. Hough, *Bioscience* **50**, 585 (2000).
22. G. W. Luck, *Biol. Rev. Camb. Philos. Soc.* **82**, 607 (2007).
23. K. K. Karanth, L. M. Curran, J. D. Reuning-Scherer, *Biol. Conserv.* **128**, 147 (2006).
24. J. S. Brashares, P. Arcese, M. K. Sam, *Proc. R. Soc. London B. Biol. Sci.* **268**, 2473 (2001).
25. A. T. Hudak, D. H. K. Fairbanks, B. H. Brockett, *Agric. Ecosyst. Environ.* **101**, 307 (2004).
26. R. DeFries, A. Hansen, A. C. Newton, M. C. Hansen, *Ecol. Appl.* **15**, 19 (2005).
27. E. W. Sanderson *et al.*, *Bioscience* **52**, 891 (2002).
28. R. DeFries, A. Hansen, B. L. Turner, R. Reid, J. G. Liu, *Ecol. Appl.* **17**, 1031 (2007).
29. Overlapping buffers in multiple PAs were combined for analysis, which decreased the sample of 306 PAs to 284 buffer areas. Countries and ecoregions with one park were excluded from statistical analysis of median buffer growth rates, although results did not vary with their inclusion.

30. Supported by NSF OISE-0502340 and the J. S. McDonnell Foundation (G.W.), the Hellman Fund (J.S.B.) and the University of California at Berkeley. We thank A. Balmford, A. Bruner, P. Coppolillo, C. Golden, C. Kremen, R. Kuriyan, J. Scharlemann, A. R. E. Sinclair, C. Stoner, two anonymous reviewers, and the Geospatial Imaging and Informatics Facility (GIIF), Museum of Vertebrate Zoology (MVZ), and Department of Environmental Science, Policy, and Management (ESPM) at the University of California at Berkeley. The authors declare no competing financial interests.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/123/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Table S1
References

9 April 2008; accepted 4 June 2008
10.1126/science.1158900

Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops

Tony Yu-Chen Tsai,[1]* Yoon Sup Choi,[1,2]* Wenzhe Ma,[3,4] Joseph R. Pomerening,[5] Chao Tang,[3,4] James E. Ferrell Jr.[1]†

A simple negative feedback loop of interacting genes or proteins has the potential to generate sustained oscillations. However, many biological oscillators also have a positive feedback loop, raising the question of what advantages the extra loop imparts. Through computational studies, we show that it is generally difficult to adjust a negative feedback oscillator's frequency without compromising its amplitude, whereas with positive-plus-negative feedback, one can achieve a widely tunable frequency and near-constant amplitude. This tunability makes the latter design suitable for biological rhythms like heartbeats and cell cycles that need to provide a constant output over a range of frequencies. Positive-plus-negative oscillators also appear to be more robust and easier to evolve, rationalizing why they are found in contexts where an adjustable frequency is unimportant.

The mammalian heart rate is normally established by the sino-atrial node. The node generates constant-amplitude action potentials at a tunable frequency of ~50 to 150 action potentials per minute, depending on the body's oxygen demands. The cell cycle oscillator may also require this combination of an adjustable frequency and invariant amplitude. The period of the cell cycle ranges from about 10 min in rapidly dividing embryos to tens of hours in rapidly dividing somatic cells (and longer in slowly dividing somatic cells), but variations in the amplitude [the peak concentration of active cyclin-dependent kinase-1 (CDK1)] of the oscillations seem neither necessary nor desirable.

Two basic types of circuits have been proposed for biological oscillators: (i) those that contain both positive and negative feedback loops and (ii) those containing only negative feedback (Table 1) (*1*–*6*). Both the sino-atrial node oscillator and the cell cycle oscillator fall into the positive-plus-negative feedback class, suggesting that this design might be better suited for generating oscillations with a tunable frequency and constant amplitude.

We tested this idea through computational studies, beginning with an ordinary differential equation model of CDK1 oscillations in the *Xenopus* embryonic cell cycle (*7*). The model includes a negative feedback loop [active CDK1 brings about its inactivation through the anaphase-promoting complex (APC)] and a pair of positive feedback loops (active CDK1 activates its activator Cdc25 and inactivates its inhibitor Wee1) (Fig. 1A). We specified the strength of the positive feedback through a parameter r, the ratio of the activities of Cdc25 and Wee1 in interphase versus M phase. Because the rate of cyclin synthesis determines the frequency of CDK1 oscillations in *Xenopus* embryos (*7*, *8*), we varied the cyclin synthesis rate constant k_{synth} in the model and determined how the amplitude and frequency of the oscillations were affected by this variation.

In the negative feedback–only version of the model ($r = 1$ in Fig. 1, B and C), a relatively small range of k_{synth} values yielded oscillations. Plotting the amplitude and frequency of the oscillations on a log-log plot yielded a tight, inverted U-shaped curve (Fig. 1B). The range of frequencies over which the oscillator functioned was small (1.7-fold), and even within this range, the frequency could not be adjusted without compromising the amplitude substantially.

Adding positive feedback markedly changed the amplitude/frequency relation (Fig. 1, B and C). At a biologically realistic feedback strength of $r = 10$ (*9*–*11*), the oscillator functioned over a 4900-fold range of frequencies (Fig. 1B, green points). Over much of this range, the frequency of the oscillator was linearly proportional to k_{synth}, and the amplitude was approximately constant (Fig. 1, B and C). Thus, positive feedback provided a highly tunable frequency and robust amplitude.

Something other than the cyclin synthesis rate may tune the frequencies of some cell cycles. We therefore asked whether the negative feedback–only oscillator might operate over a wider range of frequencies if one of the model's other 20 parameters were varied. This was not the case; invariably, the oscillator operated over only a narrow frequency range. Of course if all of the rate constants were multiplied by the same factor (equivalent to scaling the units of time), the oscillator's frequency could be varied without changing the amplitude. However, this type of coordinated regulation is not relevant to any of the biological oscillators that we are familiar with (Table 1).

[1] Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305–5174, USA. [2] School of Interdisciplinary Bioscience and Bioengineering, Pohang University of Science and Technology, Pohang, 790-784, Republic of Korea. [3] Center for Theoretical Biology, Peking University, Beijing, 100871, China. [4] California Institute for Quantitative Biosciences, University of California, San Francisco, CA 94143–2540, USA. [5] Department of Biology, Indiana University, Bloomington, IN 47405, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: james.ferrell@stanford.edu

Both the tunable frequency and constant amplitude of the positive-plus-negative feedback cell cycle model arise because the system behaves like a relaxation oscillator (*12*–*15*). Relaxation oscillators are built on a hysteretic switch, and experimental studies have shown that in *Xenopus* extracts the response of the CDK1/Cdc25/Wee1/Myt1 positive feedback loop is hysteretic, resembling that shown in Fig. 2A (*16*, *17*).

To see how relaxation oscillations can arise from a hysteretic switch and to see why this permits a tunable period and constant amplitude, assume that a cell cycle begins with no cyclin B and no active CDK1 and that cyclin B synthesis is slow relative to the phosphorylation and dephosphorylation reactions that allow the hysteretic switch to approach its steady state (Fig. 2A). As cyclin B accumulates, the system moves up the lower branch of the stimulus/response curve, and the CDK1 activity slowly rises (Fig. 2, A and B, segments 0 and 1). Ultimately, the branch terminates and the system switches ("relaxes") to the other branch (Fig. 2, A and B, segment 2).

Now assume that at this higher, mitotic level of CDK1 activity, the APC is turned on and the cyclin B concentration begins to fall slowly. The system progresses down the upper branch of the stimulus/response curve (Fig. 2, A and B, segment 3) until the branch terminates and the system switches to the lower branch (Fig. 2, A and B, segment 4). The APC turns back off, cyclin B reaccumulates, and the cycle starts over. Thus, oscillations in this system essentially represent a walk around the hysteretic steady-state stimulus/response loop. The frequency of the oscillator is determined by how rapid the walk is, and the amplitude (the height of the loop) is constant.

In reality, the rate of cyclin B destruction by the APC is not slow compared with the phosphorylation and dephosphorylation reactions (*7*). This fact is incorporated into the cell cycle model examined here, and it makes the orbits of the oscillator overshoot the hysteretic loop (Fig. 2C). Nevertheless, the model still behaves much like a relaxation oscillator, especially at low k_{synth} values (Fig. 2, C and D).

To test the generality of the idea that positive feedback enables an oscillator to have a tunable frequency and constant amplitude, we examined several other oscillator models, including five negative feedback–only models: (i) the Goodwin oscillator, a well-studied model relevant to circadian oscillations (*18*, *19*); (ii) the Repressilator, a transcriptional triple-negative feedback loop constructed in *Escherichia coli* (*20*); (iii) the "Pentilator," a Repressilator with five (rather than three) repressors; (iv) the Metabolator (*21*), a synthetic metabolic oscillator; and (v) the Frzilator, a model of the control of gliding motions in myxobacteria (*22*). In four of the cases (Goodwin, Repressilator, Pentilator, and Metabolator), the amplitude/frequency curves were inverted U-shaped curves similar to that seen for the negative feedback–only cell cycle model (Figs. 1B and 3A). In the case of the

Table 1. Positive feedback loops in biological oscillators.

Oscillator	Period	Positive feedback	Refs.
Sino-atrial pacemaker	~1 s	Depolarization → Na^{+} channel activation → depolarization	(*29*)
Calcium spikes	~100 s	Cytoplasmic Ca^{2+} → PLC → IP$_3$ → cytoplasmic Ca^{2+} Cytoplasmic Ca^{2+} → IP$_3$R → cytoplasmic Ca^{2+} Cytoplasmic Ca^{2+} → IP$_3$R -| ER Ca^{2+} -| SOC → cytoplasmic Ca^{2+}	(*25*, *30*, *31*)
Myxobacterial gliding	~10 min	None known	(*22*)
Animal cell cycle (*Xenopus laevis* embryos)	~30 min	Cdk1 → Cdc25 → Cdk1 Cdk1 -| Wee1 -| Cdk1 Cdk1 -| Myt1 -| Cdk1	(*32*, *33*)
Somitogenesis	~30 min	DeltaC → Notch → DeltaC	(*34*)
Yeast cell cycle (*S. cerevisiae*)	~2 hours	CLN1,2 transcription → CDK1 → CLN1,2 transcription CDK1 -| Sic1 -| CDK1 CDK1 -| Cdh1 -| CDK1	(*6*, *35*–*39*)
NF-κB responses	~100 min	None known	(*40*, *41*)
p53 responses	~100 min	p53 → PTEN -| Akt → Mdm2 -| p53 p53 → p21 -| Cdk2 -| Rb -| Mdm2 -| p53	(*42*, *43*)
Animal cell cycle (somatic cells)	~24 hours	CDK2 -| Rb -| E2F → CDK2 Cdk1 → Cdc25 → Cdk1 Cdk1 -| Wee1 -| Cdk1 Cdk1 -| Myt1 -| Cdk1	(*44*)
Circadian rhythm (mammals)	~24 hours	BMAL1 → Rora → BMAL1	(*45*)
Circadian rhythm (*Drosophila*)	~24 hours	CLK → PDP1 → CLK	(*45*)
Circadian rhythm (fungi)	~24 hours	FRQ → WC-1 → FRQ	(*46*)
Circadian rhythm (cyanobacteria)	~24 hours	KaiC-SP -| KaiA -| KaiC-SP	(*26*)

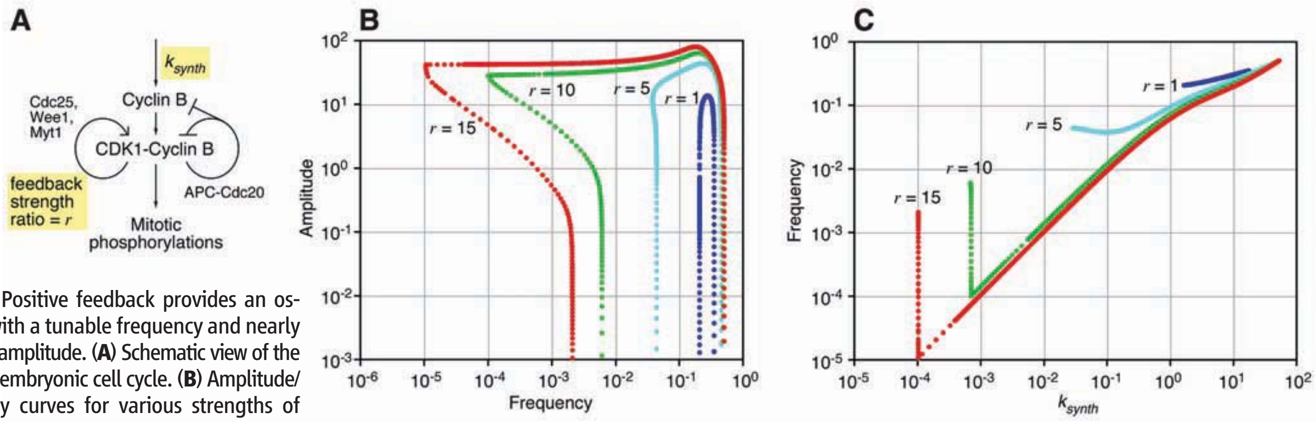


Fig. 1. Positive feedback provides an oscillator with a tunable frequency and nearly constant amplitude. (**A**) Schematic view of the *Xenopus* embryonic cell cycle. (**B**) Amplitude/frequency curves for various strengths of positive feedback (*r*). The frequency of the oscillator was changed by varying the rate constant for cyclin B synthesis, k_{synth}. (**C**) Frequency as a function of k_{synth} for various strengths of positive feedback.

Frzilator, the legs of the curve were truncated; the oscillator had a nonzero minimal amplitude (Fig. 3A). For all five of the negative feedback–only models, the oscillators functioned over only a narrow range of frequencies (Fig. 3A).

We also examined four positive-plus-negative feedback oscillators: (i) the van der Pol oscillator, inspired by studies of vacuum tubes (*12*); (ii) the Fitzhugh-Nagumo model of propagating action potentials (*23*, *24*); (iii) the Meyer-Stryer model of calcium oscillations (*25*); and (iv) a model of circadian oscillations in the cyanobacterial KaiA/B/C system (*26*–*28*). In each case, we obtained a flat, wide amplitude/frequency curve (Fig. 3B). Thus, a tunable frequency plus constant amplitude can be obtained from many different positive-plus-negative feedback models; this feature is not peculiar to one particular topology or parameterization.

These findings rationalize why the positive-plus-negative feedback design might have been selected through evolution in cases where a tunable frequency and constant amplitude are important, such as heartbeats and cell cycles. However, it is not clear that an adjustable frequency would be advantageous for circadian oscillations, because frequency is fixed at one cycle per day. Nevertheless, the cyanobacterial circadian oscillator appears to rely on positive feedback (*26*), and positive feedback loops have been postulated for other circadian oscillators as well (Table 1). This raises the question of whether the positive-plus-negative feedback design might offer additional advantages.

One possibility is that the positive-plus-negative feedback design permits oscillations over a wider range of enzyme concentrations and kinetic constant values, making the oscillator easier to evolve and more robust to variations in its imperfect components. We tested this idea through a Monte Carlo approach. We formulated three simple oscillator models: (i) a three-variable triple negative feedback loop with no additional feedback (Fig. 4A), (ii) one with added positive feedback (Fig. 4B), or (iii) one with added negative feedback (Fig. 4C). We generated random parameter sets for the models and then for each set determined whether the model produced limit cycle oscillations. We continued generating parameter sets until we had amassed 500 that gave oscillations.

For the negative feedback–only model, 500 out of 138,785 parameter sets (0.36%) yielded oscillations (Fig. 4D). For the positive-plus-negative feedback model, oscillatory parameter sets were found at a higher rate: 500 out of 23,848 parameter sets (2.1%) if we assumed weak positive feedback and 500 out of 9854 sets (5.1%) for strong positive feedback (Fig. 4D). The negative-plus-negative feedback model yielded oscillations at a lower rate than even the negative feedback–only model: 500 out of 264,672 parameter sets (0.19%) for the weaker feedback strength and 500 out of 583,263 (0.086%) for the stronger feedback. This is probably because the short negative feedback loop stabilizes the output of *A*, making it difficult for changes in *C*'s activity to be propagated onward. Thus, the positive-plus-negative feedback design was substantially more robust, by this measure, than either the negative feedback–only model or the negative-plus-negative feedback model.

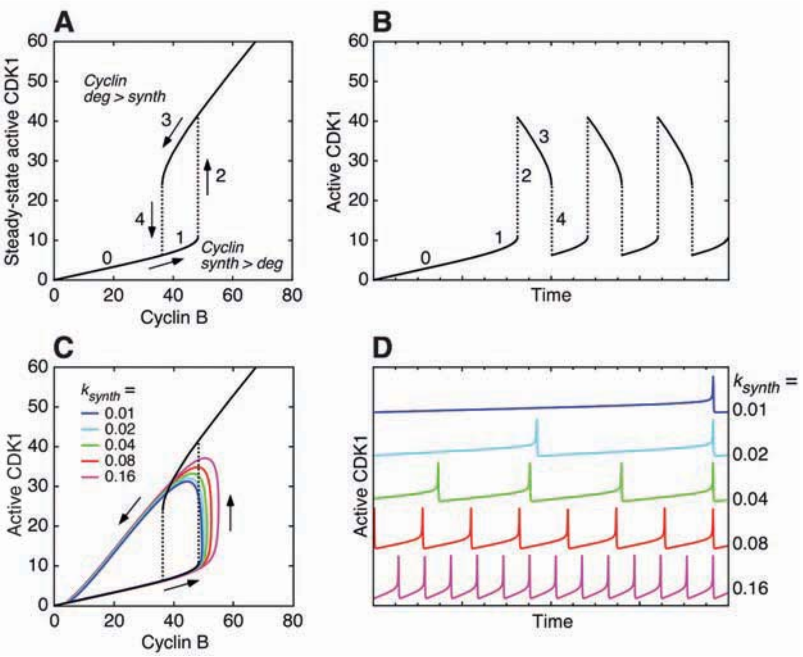


Fig. 2. From a hysteretic switch to a relaxation oscillator. (**A**) Hysteretic steady-state response of CDK1 to cyclin B, on the basis of previous experimental studies (*16*, *17*). (**B**) CDK1 activation and inactivation in the limit of slow cyclin B synthesis and degradation. (**C** and **D**) Cell cycle model run with biologically realistic parameters, showing a looser relation between the oscillations and the hysteretic steady-state response.

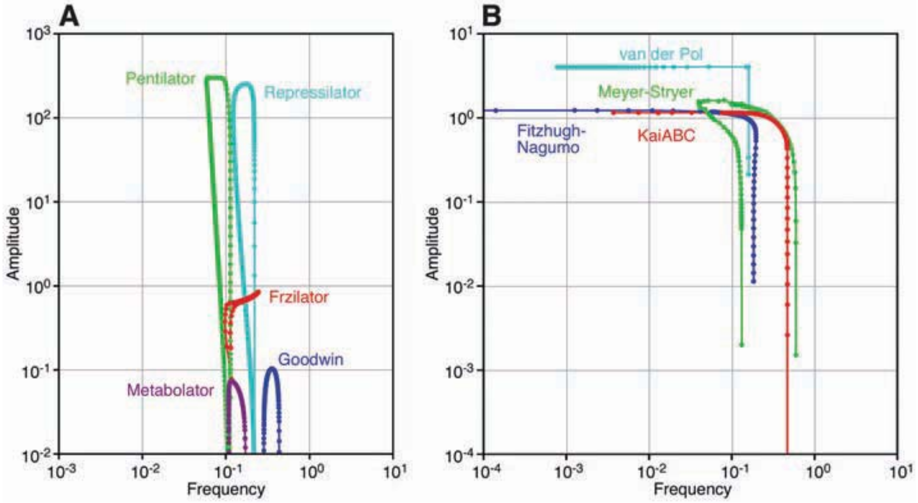


Fig. 3. Amplitude/frequency curves for various legacy oscillators. (**A**) Negative feedback–only models. (**B**) Positive-plus-negative feedback models.

The random parameter sets also provided a further test of the hypothesis that the positive-plus-negative design allows for a tunable frequency. For each oscillatory set, we varied one parameter (k_3) and calculated amplitude/frequency curves and operational frequency ranges. For the negative feedback–only and the negative-plus-negative feedback models, all of the oscillatory parameter sets yielded narrow, inverted U-shaped amplitude/frequency curves with small operational frequency ranges (Fig. 4, E and F, and fig. S1). In contrast, many of the amplitude/frequency curves for the positive-plus-negative feedback model were flat and wide, with large operational frequency ranges (Fig. 4, E and F). Thus, the positive-plus-negative design provided the possibility of a tunable frequency and near-constant amplitude.

The frequent presence of positive feedback loops in natural biological oscillators suggests

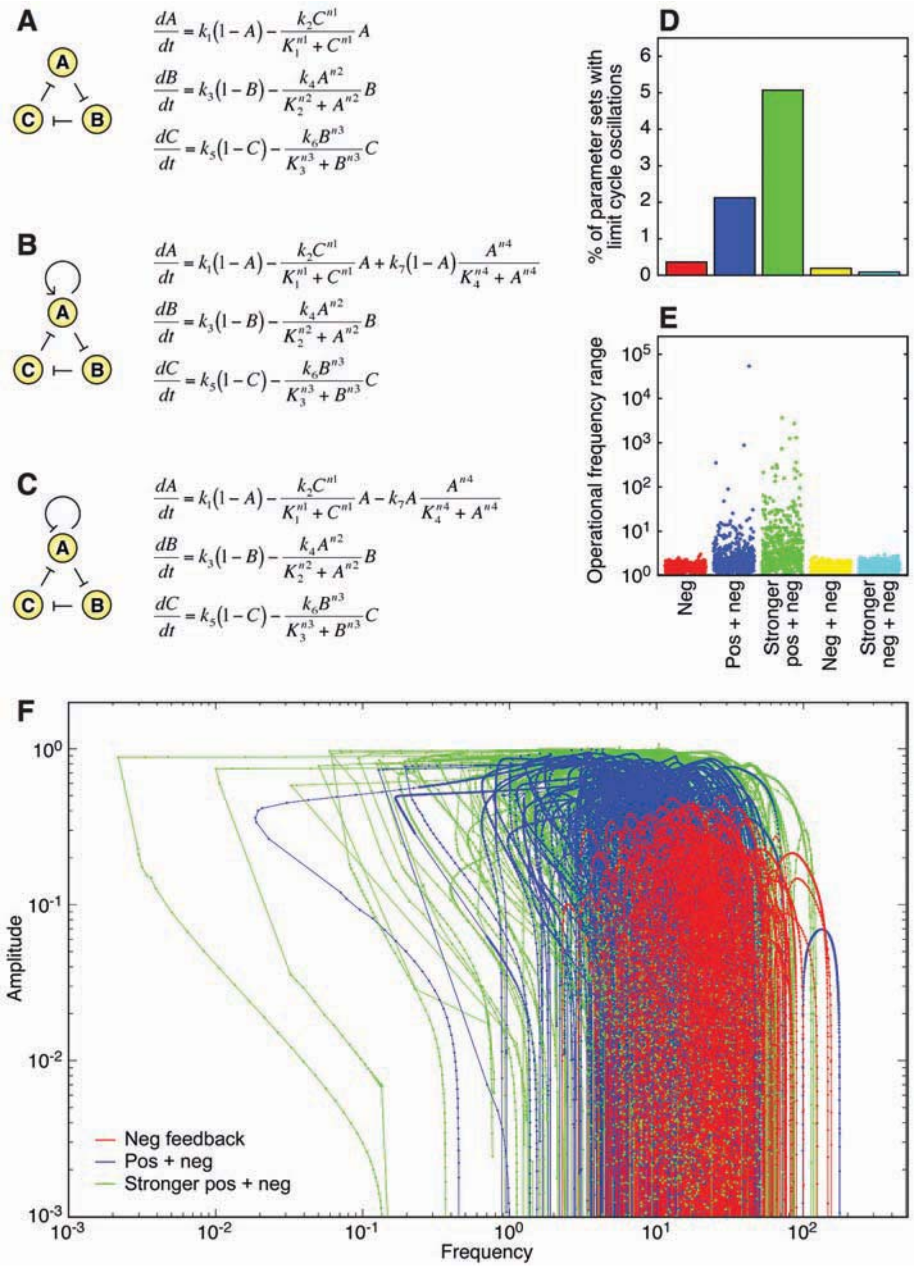


Fig. 4. Randomly parameterized oscillator models. (**A**) Negative feedback–only. *A*, *B*, and *C*, the fractions of proteins A, B, and C that are active; *K*, median effective concentration values of the Hill functions; *n*, Hill coefficients; *k*, rate constants. (**B**) Positive-plus-negative feedback. (**C**) Negative-plus-negative feedback. (**D**) Percentage of parameter sets that yielded limit cycle oscillations. For the positive-plus-negative and negative-plus-negative models, we looked at two ranges of feedback strengths: (i) $k_7 = 0$ to 100 (weak) and (ii) $k_7 = 500$ to 600 (strong). (**E**) Operational frequency ranges for the oscillators. Each point represents $freq_{\max}/freq_{\min}$ for one of the 2500 parameter sets that produced oscillations, with k_3 as the bifurcation parameter. Mean operational frequency ranges were 1.6, 370, 63, 1.6, and 1.6. Medians were 1.6, 2.2, 3.3, 1.6, and 1.6. (**F**) Amplitude/frequency curves for the randomly parameterized models. We show 300 out of 500 curves for the negative feedback–only model (red) and the positive-plus-negative feedback model with weak (blue) or strong (green) positive feedback. Curves for the negative-plus-negative feedback model are shown in fig. S1.

that this type of circuit possesses some performance advantages over simple negative feedback loops. The present work demonstrates two such advantages: (i) the ability to tune the oscillator's frequency without changing its amplitude and (ii) a greater robustness and reliability.

References and Notes

1. C. D. Thron, *Biophys. Chem.* **57**, 239 (1996).
2. A. Goldbeter *et al.*, *Chaos* **11**, 247 (2001).
3. J. J. Tyson, K. C. Chen, B. Novak, *Curr. Opin. Cell Biol.* **15**, 221 (2003).
4. P. Smolen, D. A. Baxter, J. H. Byrne, *Am. J. Physiol.* **274**, C531 (1998).
5. A. Goldbeter, *Nature* **420**, 238 (2002).
6. F. R. Cross, *Dev. Cell* **4**, 741 (2003).
7. J. R. Pomerening, S. Y. Kim, J. E. Ferrell Jr., *Cell* **122**, 565 (2005).
8. R. S. Hartley, R. E. Rempel, J. L. Maller, *Dev. Biol.* **173**, 408 (1996).
9. A. Kumagai, W. G. Dunphy, *Cell* **70**, 139 (1992).
10. P. R. Mueller, T. R. Coleman, W. G. Dunphy, *Mol. Biol. Cell* **6**, 119 (1995).
11. S. Y. Kim, E. J. Song, K. J. Lee, J. E. Ferrell Jr., *Mol. Cell. Biol.* **25**, 10580 (2005).
12. B. van der Pol, J. van der Mark, *Philos. Mag.* **6** (suppl.), 763 (1928).
13. H. S. Hahn, A. Nitzan, P. Ortoleva, J. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4067 (1974).
14. C. R. Nave, *Hyperphysics* (1995), (http://hyperphysics.phy-astr.gsu.edu/Hbase/electronic/relaxo.html).
15. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Westview, Cambridge MA, 1994).
16. W. Sha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 975 (2003).
17. J. R. Pomerening, E. D. Sontag, J. E. Ferrell Jr., *Nat. Cell Biol.* **5**, 346 (2003).
18. B. C. Goodwin, Ed. *Oscillatory Behavior in Enzymatic Control Processes*, vol. 3 (Permagon, Oxford, 1965), pp. 425–438.
19. P. Ruoff, S. Mohsenzadeh, L. Rensing, *Naturwissenschaften* **83**, 514 (1996).
20. M. B. Elowitz, S. Leibler, *Nature* **403**, 335 (2000).
21. E. Fung *et al.*, *Nature* **435**, 118 (2005).
22. O. A. Igoshin, A. Goldbeter, D. Kaiser, G. Oster, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15760 (2004).
23. R. FitzHugh, *Biophys. J.* **1**, 445 (1961).
24. J. Nagumo, S. Arimoto, S. Yoshizawa, *Proc. IRE* **50**, 2061 (1964).
25. T. Meyer, L. Stryer, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5051 (1988).
26. M. J. Rust, J. S. Markson, W. S. Lane, D. S. Fisher, E. K. O'Shea, *Science* **318**, 809 (2007); published online 4 October 2007, 10.1126/science.1148596.
27. M. Nakajima *et al.*, *Science* **308**, 414 (2005).
28. T. Nishiwaki *et al.*, *EMBO J.* **26**, 4029 (2007).
29. A. L. Hodgkin, A. F. Huxley, *J. Physiol.* **117**, 500 (1952).
30. R. S. Lewis, *Annu. Rev. Immunol.* **19**, 497 (2001).
31. M. J. Berridge, *Novartis Found. Symp.* **239**, 52 (2001).
32. B. Novak, J. J. Tyson, *J. Cell Sci.* **106**, 1153 (1993).
33. J. J. Tyson, B. Novak, *J. Theor. Biol.* **210**, 249 (2001).
34. A. Mara, S. A. Holley, *Trends Cell Biol.* **17**, 593 (2007).
35. K. Levine, A. H. Tinkelenberg, F. Cross, *Prog. Cell Cycle Res.* **1**, 101 (1995).
36. J. M. Bean, E. D. Siggia, F. R. Cross, *Mol. Cell* **21**, 3 (2006).
37. D. Stuart, C. Wittenberg, *Genes Dev.* **9**, 2780 (1995).
38. L. Dirick, K. Nasmyth, *Nature* **351**, 754 (1991).
39. K. C. Chen *et al.*, *Mol. Biol. Cell* **15**, 3841 (2004).
40. A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, *Science* **298**, 1241 (2002).
41. D. E. Nelson *et al.*, *Science* **306**, 704 (2004).
42. R. Lev Bar-Or *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11250 (2000).
43. G. Lahav *et al.*, *Nat. Genet.* **36**, 147 (2004).
44. L. A. Buttitta, B. A. Edgar, *Curr. Opin. Cell Biol.* **19**, 697 (2007).
45. M. Gallego, D. M. Virshup, *Nat. Rev. Mol. Cell Biol.* **8**, 139 (2007).
46. K. Lee, J. J. Loros, J. C. Dunlap, *Science* **289**, 107 (2000).
47. We thank E. Sontag, A. Millar, and B. Novak for helpful discussions; J. Hasty for communicating unpublished results; and J. Ubersax and G. Anderson for comments on the manuscript. This work was supported by grants from NIH (GM61726 and GM77544), by the Li Foundation, and by a Korea Science and Engineering Foundation grant from the Korean government (No. R15-2004-033-05002-0).

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/126/DC1
SOM Text
Fig. S1
Table S1
References

25 February 2008; accepted 9 June 2008
10.1126/science.1156951

Sporadic Autonomic Dysregulation and Death Associated with Excessive Serotonin Autoinhibition

Enrica Audero,[1] Elisabetta Coppi,[2] Boris Mlinar,[2] Tiziana Rossetti,[1] Antonio Caprioli,[3] Mumna Al Banchaabouchi,[1] Renato Corradetti,[2] Cornelius Gross[1]*

Sudden infant death syndrome is the leading cause of death in the postneonatal period in developed countries. Postmortem studies show alterations in serotonin neurons in the brainstem of such infants. However, the mechanism by which altered serotonin homeostasis might cause sudden death is unknown. We investigated the consequences of altering the autoinhibitory capacity of serotonin neurons with the reversible overexpression of serotonin 1A autoreceptors in transgenic mice. Overexpressing mice exhibited sporadic bradycardia and hypothermia that occurred during a limited developmental period and frequently progressed to death. Moreover, overexpressing mice failed to activate autonomic target organs in response to environmental challenges. These findings show that excessive serotonin autoinhibition is a risk factor for catastrophic autonomic dysregulation and provide a mechanism for a role of altered serotonin homeostasis in sudden infant death syndrome.

Sudden infant death syndrome (SIDS) is the leading cause of postneonatal death in the United States (*1*). Postmortem studies have reported robust morphological and biochemical deficits in medullary raphe serotonin neurons in SIDS brains (*2*, *3*). Immunohistochemical analysis revealed an increased number of serotonin neurons, as well as an increase in the fraction of serotonin neurons showing an immature, granular cell morphology (*3*), which suggests a failure or delay in the maturation of these neurons in SIDS infants. Furthermore, a significant decrease in serotonin receptor 1A (Htr1a) and decrease in relative serotonin transporter–binding density were observed. These findings point to a deficit in serotonin function as a risk factor for SIDS. However, until now, no serotonin deficits capable of causing a catastrophic dysregulation of autonomic circuits have been identified.

In mammals, normal levels of serotonin neuron firing are actively maintained by negative feedback inhibition via the somatodendritic autoreceptor Htr1a (*4*–*8*). Pharmacological activation of Htr1a receptors in the medullary raphe reduces serotonin neuron firing and decreases heart rate, body temperature, and respiration (*9*–*11*). We examined the effects of increasing negative feedback inhibition of serotonin by the reversible and tissue-specific overexpression of Htr1a autoreceptors in transgenic mice. Mice with conditional overexpression of Htr1a in serotonin neurons of the raphe nuclei were produced by driving expression of Htr1a under the control of the serotonin transporter gene (5-HTT, or *Slc6a4*) promoter using the tetracycline-off (tet-OFF) system (*12*, *13*) (Fig. 1A). Mice carrying the gene encoding the tetracycline transactivator (tTA) under the control of the endogenous serotonin transporter promoter (*Slc6a4*^{tTA}) were crossed with mice carrying a tTA-inducible allele of the

[1] Mouse Biology Unit, European Molecular Biology Laboratory (EMBL), Via Ramarini 32, 00015 Monterotondo, Italy. [2] Department of Preclinical and Clinical Pharmacology, University of Firenze, Viale Giacoro Pieraccini 6, 50139 Firenze, Italy. [3] Laboratory of Behavioural Neuropharmacology, Sigma-Tau SpA, Via Pontina Kilometer 30.400, 00040 Pomezia, Italy.

*To whom correspondence should be addressed. E-mail: gross@embl.it

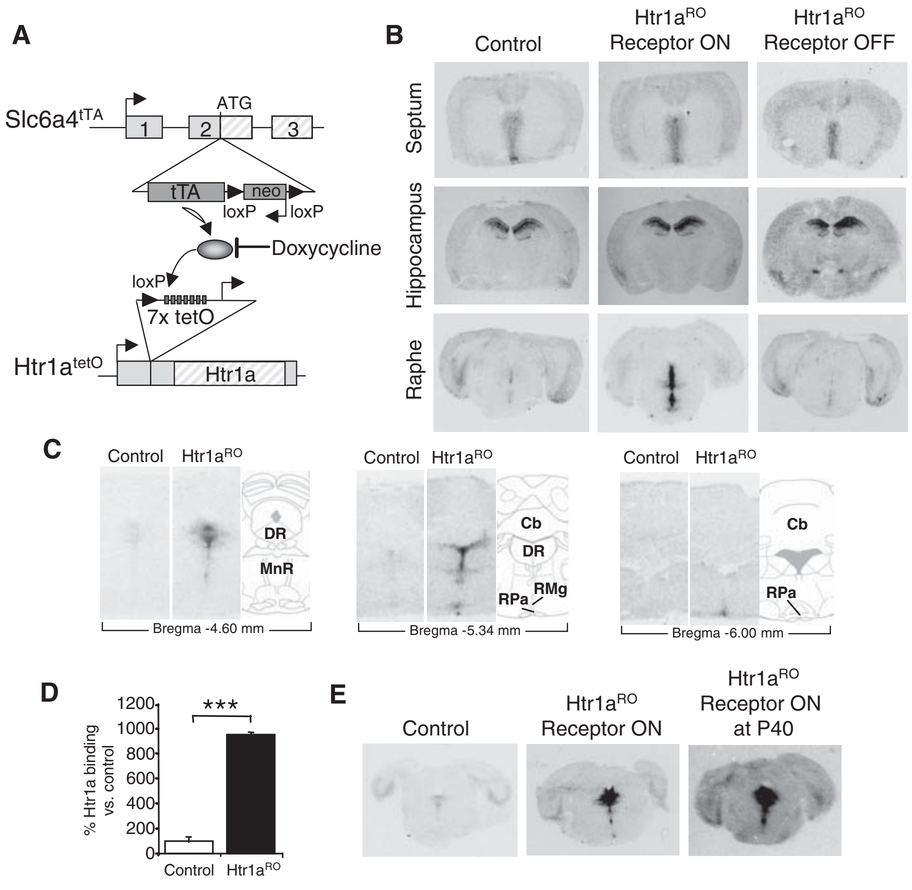


Fig. 1. Transgenic mice with reversible overexpression of Htr1a in serotonin neurons. (**A**) Mice in which tTA-binding sites (7x tetO) and a minimal cytomegalovirus promoter were engineered into the 5′ untranslated region of Htr1a (Htr1a^{tetO}) were crossed with mice in which the tTA coding sequence was inserted at the start codon of the serotonin transporter (Slc6a4^{tTA}). (**B**) Htr1a^{RO} mice showed doxycycline-sensitive overexpression of Htr1a protein exclusively in raphe nuclei, as shown by ^{125}I-MPPI–4-(2′-Methoxyphenyl)-1-[2′-(*n*-2′′-pyridinyl)-*p*-[^{125}I]iodobenzamido]ethylpiperazine) autoradiography, when compared with littermate control mice (P21, *n* = 2 mice). (**C**) Htr1a^{RO} mice show overexpression of Htr1a protein in the dorsal raphe (DR), median raphe (MnR), raphe magnus (RMg), and raphe pallidus (Rpa). Cb, cerebellum. (**D**) Quantification of Htr1a protein levels in untreated control and Htr1a^{RO} mice shown in (E) (*n* = 4 mice; *** $P < 0.001$). (**E**) Htr1a protein expression was induced in P60 Htr1a^{RO} mice upon removal of doxycycline at P40 (receptor ON at P40) to levels similar to those seen in untreated mice (receptor ON).

endogenous gene encoding Htr1a (*Htr1a*^{tetO}) (*13*). Double transgenic mice (Htr1a raphe-overexpressing mice, or Htr1a^{RO}) showed Htr1a protein overexpression in raphe nuclei of the mid- and hindbrain (Fig. 1, B and C). Quantitative autoradiographic analysis revealed approximately 10-fold more receptor protein in Htr1a^{RO} mice as compared with control littermates (Fig. 1D). Receptor overexpression could be reversed by treatment with doxycycline (Fig. 1B). Removal of doxycycline after treatment until postnatal day 40 (P40) was able to induce Htr1a overexpression in adult animals (Fig. 1E).

To examine changes in serotonin autoinhibition, we performed extracellular single-unit recordings from serotonin neurons in brain slices of Htr1a^{RO} and control mice (*14*). Addition of the serotonin precursor tryptophan to the brain slice increased endogenous serotonin levels (*15*, *16*) and consequently caused a reduction in cell firing. In both genotypes, this effect was reversed by the selective Htr1a antagonist WAY100635 (Fig. 2, A and D). Application of tryptophan to brain slices from Htr1a^{RO} mice caused significantly greater reduction in cell firing than that seen in slices from control mice (Fig. 2, B and C). Htr1a^{RO} mice treated with doxycycline showed autoinhibition similar to that in control mice (Fig. 2C). Increased serotonin autoinhibition would be expected to depress ongoing serotonin neurotransmission in the intact animal. Accordingly, the ratio of the serotonin metabolite, 5-hydroxyindole acetic acid, to serotonin (5-HIAA/5-HT) was significantly decreased in Htr1a^{RO} mice as compared with control mice (Fig. 2E). This effect was absent in mice treated with doxycycline and could be induced after the removal of doxycycline at P40 (Fig. 2, F and G).

Pharmacological depletion of serotonin in the brain of humans and other mammals is associated with changes in physiology and behavior (*17*). Nevertheless, mice with a selective genetic deletion of serotonin neurons are viable (*18*, *19*), as are mice in which all serotonin receptors and the serotonin plasma membrane and vesicular transporters have been knocked out [although see (*20*, *21*)]. Thus, it came as a surprise that a majority of Htr1a^{RO} mice died before reaching 3 months of age (Fig. 3, A and B). Death was most frequent between P25 and P80, and only approximately 30% of Htr1a^{RO} mice survived beyond P120. Death could be prevented by treating Htr1a^{RO} animals continuously with doxycycline (Fig. 3, A and B). Death could also be prevented by implanting Htr1a^{RO} mice with osmotic minipumps continuously delivering WAY100635 (treatment from P30 to P58 with the vehicle resulted in 28.5% death; with WAY100635, no deaths). The observation that a substantial fraction of animals survived into late adulthood suggested that either Htr1a^{RO} mice were more susceptible to death during a limited developmental time window, or alternatively, that genetic or environmental factors imparted resistance to a subset of animals. We therefore collected survival data for mice in which Htr1a was induced at different developmental time points. Overexpression of Htr1a beginning at P60 led to significantly fewer deaths than overexpression beginning at P40 (30 versus

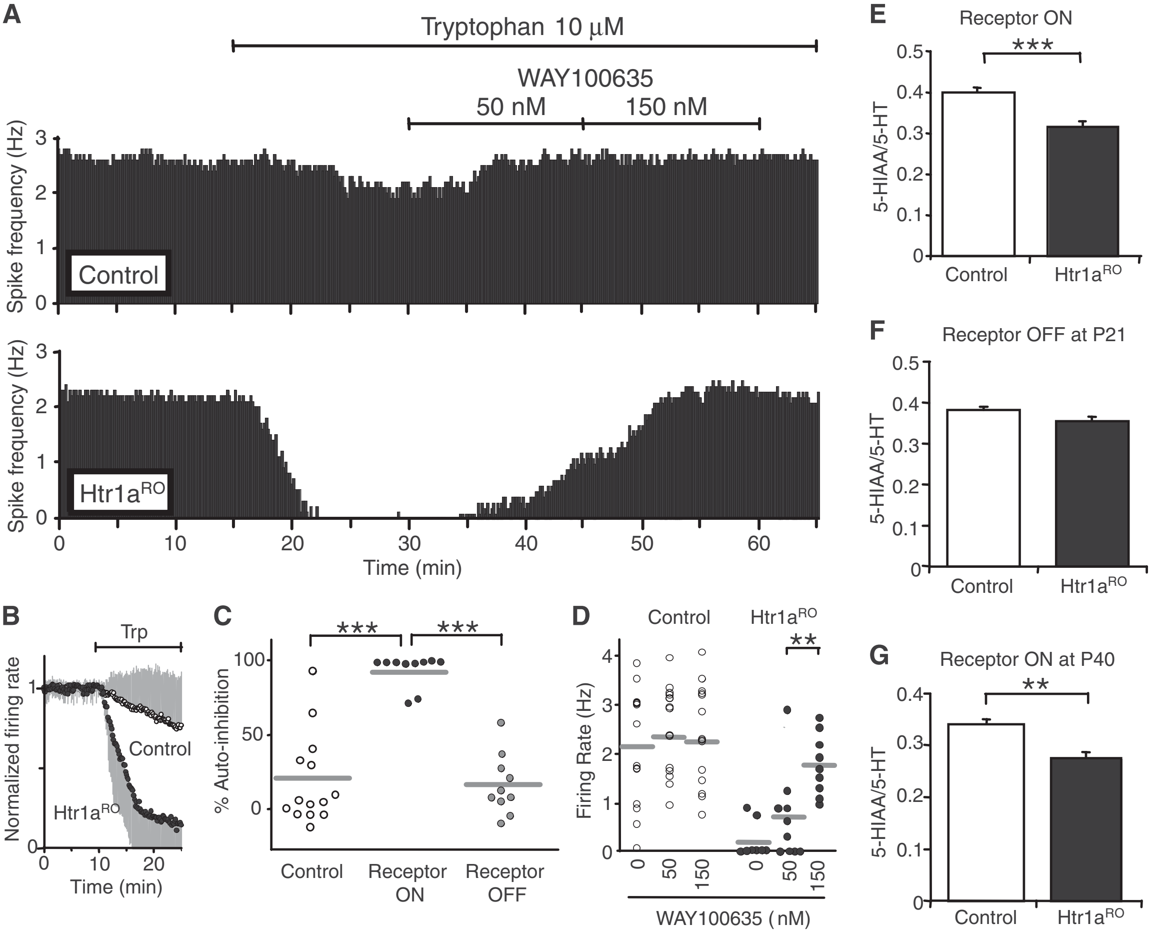


Fig. 2. Increased autoinhibition of serotonin cells and decreased serotonin turnover in Htr1a^{RO} mice. (**A**) Time course of firing rate from two representative serotonin cells recorded from the dorsal raphe nucleus in brain slices of control and Htr1a^{RO} mice. Application of the serotonin precursor tryptophan to the slice produced a greater reduction in firing rate in Htr1a^{RO} mice than in controls. Application of the selective Htr1a receptor antagonist WAY100635 reversed the effect of tryptophan, revealing that autoinhibition is mediated by Htr1a receptors. (**B**) Summary time course of tryptophan (Trp) (10 μM) effect in control and Htr1a^{RO} mice (n = 15 control cells; n = 11 Htr1a^{RO} cells; mean ± SD). (**C**) The percentage of autoinhibition in individual serotonin neurons was significantly greater in Htr1a^{RO} cells (receptor ON, 94.2 ± 11.7%; n = 10 cells; mean ± SD) when compared with cells from control animals (19.1 ± 30.7%; n = 14 cells; mean ± SD) and to cells from Htr1a^{RO} animals treated continuously with doxycycline (receptor OFF, 13.4 ± 17.2%; n = 10 cells; mean ± SD; *** $P < 0.0001$; two-tailed Mann-Whitney test). (**D**) Scatterplot showing a significant increase in the firing rate of cells from Htr1a^{RO} mice after application of WAY100635 in the presence of tryptophan (** $P < 0.01$; two-tailed Wilcoxon test). (**E**) 5-HIAA/5-HT ratio, as measured by high-performance liquid chromatography analysis of whole-brain tissue extracts from adult mice, was significantly decreased in Htr1a^{RO} mice as compared with control littermates (n = 12 mice). Deficits in serotonin turnover were (**F**) normalized by doxycycline treatment initiated at P21 (n = 5 to 7 mice) and (**G**) restored after doxycycline removal at P40 (n = 8 mice) when compared with similarly treated control littermates (*** $P < 0.001$, ** $P < 0.01$).

90% after 120 days, $P = 0.024$; Fig. 3, C and D), confirming the existence of a sensitive period for the death phenotype.

Activation of Htr1a autoreceptors suppresses heart rate and body temperature (*10*, *11*). We thus performed simultaneous radiotelemetric monitoring of cardiac activity [electrocardiogram (ECG)], body temperature, and locomotor activity in Htr1a^{RO} and littermate control mice. Baseline heart rate and body temperature were normal in Htr1a^{RO} animals, consistent with robust homeostatic control of autonomic activity, despite altered serotonin neurotransmission in these mice (fig. S1). However, Htr1a^{RO} mice showed sporadic autonomic crises characterized by decreases in heart rate and body temperature that persisted for several hours and frequently recovered only after several days (Fig. 4, A to D). No crises were observed in control littermate animals. Seventy-three percent of the animals (11/15) showed at least one crisis, and in 37% of the observed crises (7/19) severe bradycardia and hypothermia were irreversible and progressed to death (Fig. 4, C and D). Finally, detailed analysis of autonomic crises in Htr1a^{RO} mice revealed that bradycardia preceded hypothermia by 5 to 10 min at the initiation of a crisis (fig. S2), consistent with parallel alterations in multiple autonomic circuits.

The circuits mediating the serotonergic maintenance of autonomic activity primarily involve descending projections emanating from the medullary raphe nuclei. Serotonin neurons in these nuclei enervate the intermediolateral cell column of the spinal cord that contains sympathetic preganglionic neurons that positively modulate critical autonomic target organs such as the heart and brown adipose tissue (*22*–*24*). To determine whether defective activation of downstream autonomic circuits could underlie the sporadic death seen in Htr1a^{RO} mice, we examined brown adipose tissue activation in response to cold stress. Exposure to cold selectively increases the firing of medullary serotonin neurons (*25*) and is associated with an increased expression of mitochondrial uncoupling protein 1 (UCP1) in brown adipose tissue. UCP1 is responsible for diverting energy from the production of adenosine 5′-triphosphate to the generation of heat for the maintenance of thermostasis (*26*, *27*). Although exposure for 30 min to 4°C caused a significant increase in UCP1 mRNA levels in brown adi-

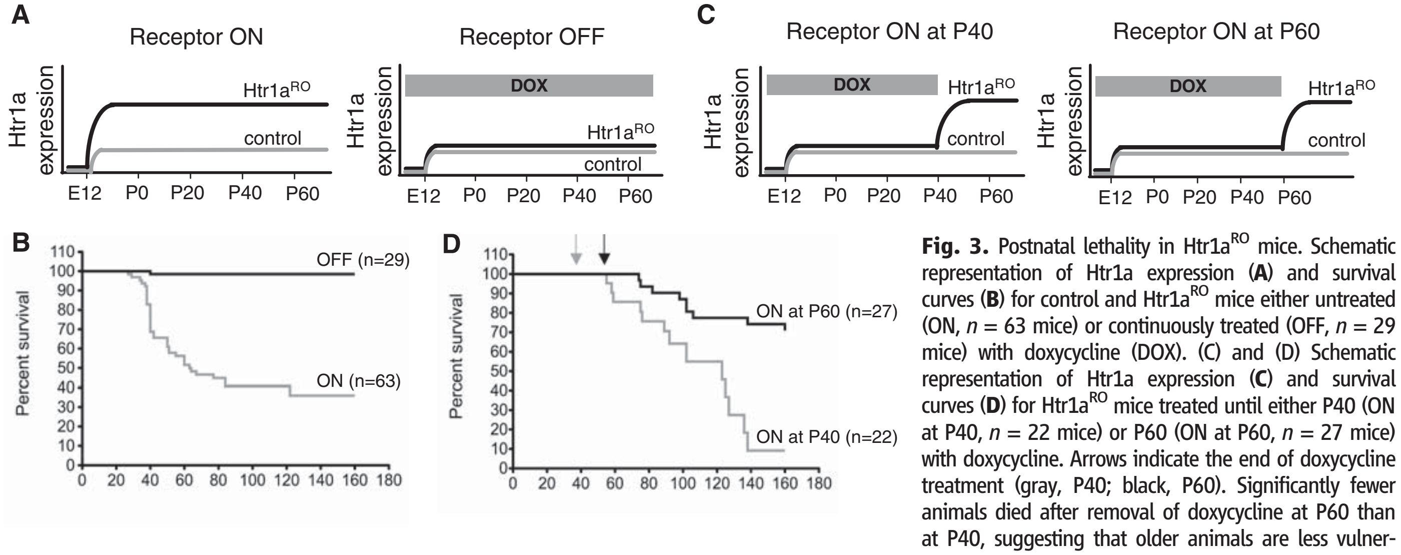


Fig. 3. Postnatal lethality in Htr1a^{RO} mice. Schematic representation of Htr1a expression (**A**) and survival curves (**B**) for control and Htr1a^{RO} mice either untreated (ON, n = 63 mice) or continuously treated (OFF, n = 29 mice) with doxycycline (DOX). (C) and (D) Schematic representation of Htr1a expression (**C**) and survival curves (**D**) for Htr1a^{RO} mice treated until either P40 (ON at P40, n = 22 mice) or P60 (ON at P60, n = 27 mice) with doxycycline. Arrows indicate the end of doxycycline treatment (gray, P40; black, P60). Significantly fewer animals died after removal of doxycycline at P60 than at P40, suggesting that older animals are less vulnerable to Htr1a overexpression (P = 0.024).

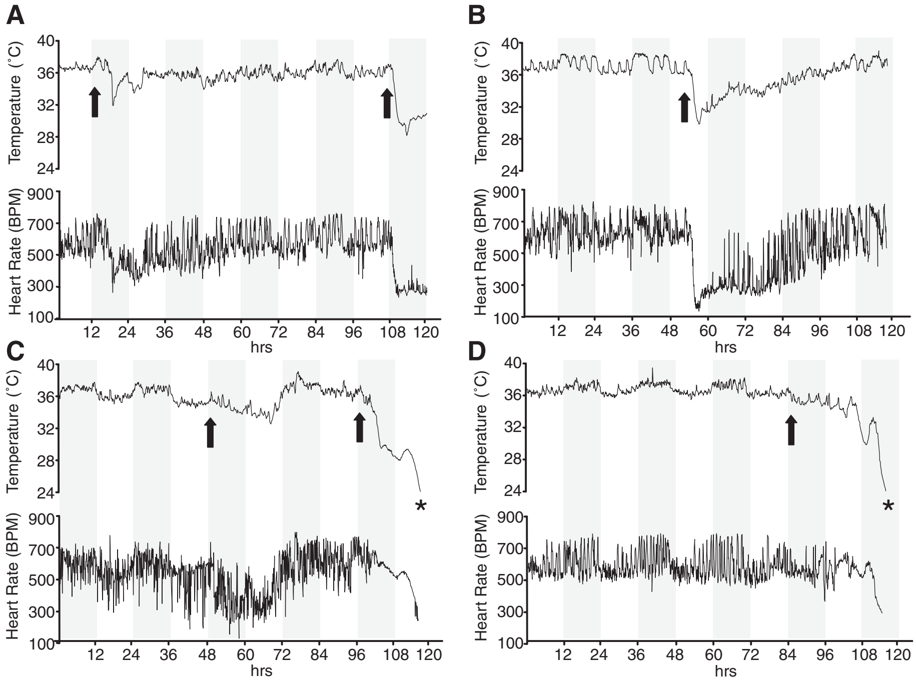


Fig. 4. Sporadic autonomic crises in Htr1a^{RO} mice. (**A** to **D**) Continuously recorded radiotelemetric assessment of body temperature and heart rate for four mice showing autonomic crises. Animals were treated with doxycycline until P60, transponders for radiotelemetry recordings were implanted, and the antibiotic was then removed 1 week after surgery. Crises (arrows) were characterized by hypothermia and bradycardia that in some cases resolved [(A) and (B)] and in other cases progressed to death [(C) and (D); * = death). Gray bars indicate dark periods.

pose tissue from control mice, activation of UCP1 expression was completely absent in brown adipose tissue from Htr1a^{RO} animals (fig. S3). Htr1a^{RO} mice showed a small but significant decrease in UCP1 mRNA levels after cold stress (fig. S3). The origin of this paradoxical response is not clear, but may reflect a loss of serotonergic tone after excessive autoinhibition.

Although technical limitations prevented us from directly measuring the firing rate of serotonin neurons in our animals during autonomic crises, it is likely that crises were associated with a suppression of serotonin neuronal firing due to elevated autoinhibition. The events that precipitated crises in our animals are not known, and thus far we have not been able to identify environmental stressors that induce crises. However, we speculate that crises may occur preferentially after rapid changes in serotonin neuron activity, such as during sleep-wake transitions when serotonin neuron–firing rates rapidly increase and sympathetic activity must keep pace with locomotor arousal (*28*). Although crises often took place during the beginning of the wake period of the circadian cycle (Fig. 4), further studies using simultaneous measurements of electroencephalography, electromyography, ECG, and temperature in alert mice are needed to determine whether crises were strictly associated with sleep state.

The parallel decrease in heart rate and body temperature we observed during crises suggests that these events are associated with a general failure to maintain proper sympathetic tone. Such a deficit in sympathetic activation is confirmed by the complete absence of brown adipose tissue UCP1 induction in Htr1a^{RO} mice in response to cold stress (fig. S3). Heart rate and respiration data obtained from a small number of infants monitored during SIDS events revealed a pronounced bradycardia that preceded apnea in all cases (*29*). Furthermore, at least one prospective study of neonates revealed an increase in QT interval (the measure of the time between the start of the **Q wave** and the end of the **T wave** in the heart's electrical cycle), an electrocardiogram feature associated with reduced sympathetic tone, in some infants who later died of SIDS (*30*). Thus, autonomic crises in our animals appear to share critical features with SIDS.

The triple-risk model of SIDS (*31*) argues for the simultaneous occurrence of an intrinsic susceptibility, an exogenous stressor, and a critical developmental time period. Our findings show that altered serotonin homeostasis alone is sufficient to precipitate catastrophic autonomic failure and death in the absence of overt stressors. Although it appears that SIDS infants do not exhibit increased Htr1a autoreceptor expression (*3*), it is possible that they harbor functionally equivalent deficits in serotonin homeostasis that include alterations in local serotonin release, changes in intrinsic electrophysiological properties of serotonin neurons, and deficiencies in autoregulatory feedback networks, such as those involving noradrenaline (*32*) or γ-aminobutyric acid (GABA) (*33*). Moreover, although there does not seem to be a simple correspondence between the susceptible period defined in our mice (P25 to P80) and that known for SIDS (birth to 1 year), comparative anatomical studies position the mouse equivalent of human birth within the third postnatal week (*34*), consistent with the beginning of the susceptible period in our animals. However, certain features of SIDS are not modeled in our animals, including sex differences in the incidence of death (*35*) and Htr1a-binding density (*3*). Our findings link deficient serotonin homeostasis to sporadic autonomic crisis and sudden death and suggest that Htr1a^{RO} mice can serve as an animal model to help identify diagnostic and prophylactic avenues for the prevention of SIDS.

References and Notes

1. M. Heron, *Natl. Vital Stat. Rep.* **56**, 1 (2007).
2. H. C. Kinney *et al.*, *J. Neuropathol. Exp. Neurol.* **62**, 1178 (2003).
3. D. S. Paterson *et al.*, *JAMA* **296**, 2124 (2006).
4. D. J. McGinty, R. M. Harper, *Brain Res.* **101**, 569 (1976).
5. C. A. Fornal *et al.*, *J. Pharmacol. Exp. Ther.* **270**, 1345 (1994).
6. R. Guzman-Marin *et al.*, *Brain Res.* **875**, 23 (2000).
7. M. Richer, R. Hen, P. Blier, *Eur. J. Pharmacol.* **435**, 195 (2002).
8. N. Urbain, K. Creamer, G. Debonnel, *J. Physiol.* **573**, 679 (2006).
9. N. C. Taylor, A. Li, E. E. Nattie, *J. Physiol.* **566**, 543 (2005).
10. Y. Ootsuka, W. W. Blessing, *Neurosci. Lett.* **395**, 170 (2006).
11. K. Nakamura, S. F. Morrison, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **292**, R127 (2007).
12. A. Kistner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10933 (1996).
13. C. Gross *et al.*, *Nature* **416**, 396 (2002).
14. Materials and methods are available as supporting material on *Science* Online.
15. R. J. Liu, E. K. Lambe, G. K. Aghajanian, *Eur. J. Neurosci.* **21**, 945 (2005).
16. B. Mlinar *et al.*, *Neuroreport* **16**, 1351 (2005).
17. I. Lucki, *Biol. Psychiatry* **44**, 151 (1998).
18. T. J. Hendricks *et al.*, *Neuron* **37**, 233 (2003).
19. Z. Q. Zhao *et al.*, *J. Neurosci.* **26**, 12781 (2006).
20. C. G. Nebigil *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9508 (2000).
21. J. T. Erickson, G. Shafer, M. D. Rossetti, C. G. Wilson, E. S. Deneris, *Respir. Physiol. Neurobiol.* **159**, 85 (2007).
22. S. F. Morrison, *Am. J. Physiol.* **276**, R962 (1999).
23. P. Mason, *Annu. Rev. Neurosci.* **24**, 737 (2001).
24. K. Nakamura *et al.*, *J. Neurosci.* **24**, 5370 (2004).
25. F. J. Martin-Cora, C. A. Fornal, C. W. Metzler, B. L. Jacobs, *Neuroscience* **98**, 301 (2000).
26. S. Enerback *et al.*, *Nature* **387**, 90 (1997).
27. V. Golozoubova *et al.*, *FASEB J.* **15**, 2048 (2001).
28. B. L. Jacobs, F. J. Martin-Cora, C. A. Fornal, *Brain Res. Brain Res. Rev.* **40**, 45 (2002).
29. R. G. Meny, J. L. Carroll, M. T. Carbone, D. H. Kelly, *Pediatrics* **93**, 44 (1994).
30. P. J. Schwartz *et al.*, *J. Med.* **338**, 1709 (1998).
31. J. J. Filiano, H. C. Kinney, *Biol. Neonate* **65**, 194 (1994).
32. N. Haddjeri, N. Lavoie, P. Blier, *Neuropsychopharmacology* **29**, 1800 (2004).
33. P. Celada, M. V. Puig, J. M. Casanovas, G. Guillazo, F. Artigas, *J. Neurosci.* **21**, 9917 (2001).
34. B. Clancy, R. B. Darlington, B. L. Finlay, *Neuroscience* **105**, 7 (2001).
35. Task Force on Sudden Infant Death Syndrome, *Pediatrics* **116**, 1245 (2005).
36. We thank J. Gingrich for providing portions of the targeting vector for the serotonin transporter, X.-X. Zhuang for support in embryonic stem cell handling, M.-M. Zhou for help with the initial characterization of the *Slc6a4*^{tTA} line, F. Zonfrillo and S. Santanelli for animal husbandry, A.-C. Trillat for critical experimental advice, and W. Fifer and M. Myers for helpful discussions on the relevance of our model to SIDS. This work was supported by grants from NIH (MH64948, C.G.), the **N**ational **A**lliance for **R**esearch on **S**chizophrenia and **D**epression (C.G.), the Fritz Thyssen Foundation (C.G.), European Commission (LSHM-CT-2004-503474) (R.C.), and funds from EMBL (E.A., M.A.B., C.G.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/321/5885/130/DC1
Materials and Methods
Figs. S1 to S3
References

17 March 2008; accepted 5 June 2008
10.1126/science.1157871

Myosin I Can Act As a Molecular Force Sensor

Joseph M. Laakso, John H. Lewis, Henry Shuman, E. Michael Ostap*

The ability to sense molecular tension is crucial for a wide array of cellular processes, including the detection of auditory stimuli, control of cell shape, and internalization and transport of membranes. We show that myosin I, a motor protein that has been implicated in powering key steps in these processes, dramatically alters its motile properties in response to tension. We measured the displacement generated by single myosin I molecules, and we determined the actin-attachment kinetics with varying tensions using an optical trap. The rate of myosin I detachment from actin decreases >75-fold under tension of 2 piconewtons or less, resulting in myosin I transitioning from a low (<0.2) to a high (>0.9) duty-ratio motor. This impressive tension sensitivity supports a role for myosin I as a molecular force sensor.

Myosin I's are the widely expressed, single-headed, and membrane-associated members of the myosin superfamily that participate in regulating membrane dynamics and structure in nearly all eukaryotic cells. Eight myosin I isoforms are expressed in humans, making it the largest "unconventional" myosin family (*1*). One specific and well-characterized molecular function of a myosin I isoform (myo1c) is to dynamically provide tension to sensitize mechanosensitive ion channels responsible for hearing (*2*–*4*). Myosin I's also power the transport and deformation of membranes in the cell cortex and in apical cell projections (*5*–*8*). To perform these roles, myosin I's have been proposed to act as tension-sensing proteins that alter their adenosine triphosphatase (ATPase) and mechanical properties in response to changes in loads imparted by their cellular cargos (*3*, *9*).

Biochemical, structural, and single-molecule experiments suggest that some myosin I isoforms (myo1a, myo1b, and myo1c) are adapted to sense tension. Specifically, it has been shown that myosin I produces its working stroke displacement in two substeps (*10*). An initial displacement of the lever arm is followed by an additional ~32° rotation that accompanies adenosine diphosphate (ADP) release (*11*). Because ADP release kinetically limits the rate of the detachment of myosin from actin (*12*, *13*), the extra lever-arm rotation has been proposed to be a force-sensing substep, with the presence of resisting loads preventing this lever-arm rotation and thus inhibiting ADP release and actin detachment. A similar model has been proposed for gating of myosin V motor activity during processive motility (*14*).

We characterized the motor activity of myosin I by measuring single-molecule force-generating events using the three-bead configuration, in which a single actin filament, suspended between two beads held by separate optical traps, is brought close to the surface of a pedestal bead that is sparsely coated with myosin (*15*). A recombinant myo1b splice isoform containing five calmodulin-binding IQ motifs and a C-terminal biotinylation tag (*16*) was attached to streptavidin-coated pedestal beads (*17*). Single-molecule actomyosin interactions at low loads were acquired using low trap stiffness (~0.022 pN/nm, Fig. 1A) in the presence of 1 to 50 μM ATP. Because the myo1b working

The Pennsylvania Muscle Institute and Department of Physiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA.

*To whom correspondence should be addressed at the Department of Physiology, University of Pennsylvania School of Medicine, B400 Richards Building, Philadelphia, PA 19104–6085, USA. E-mail: ostap@mail.med.upenn.edu

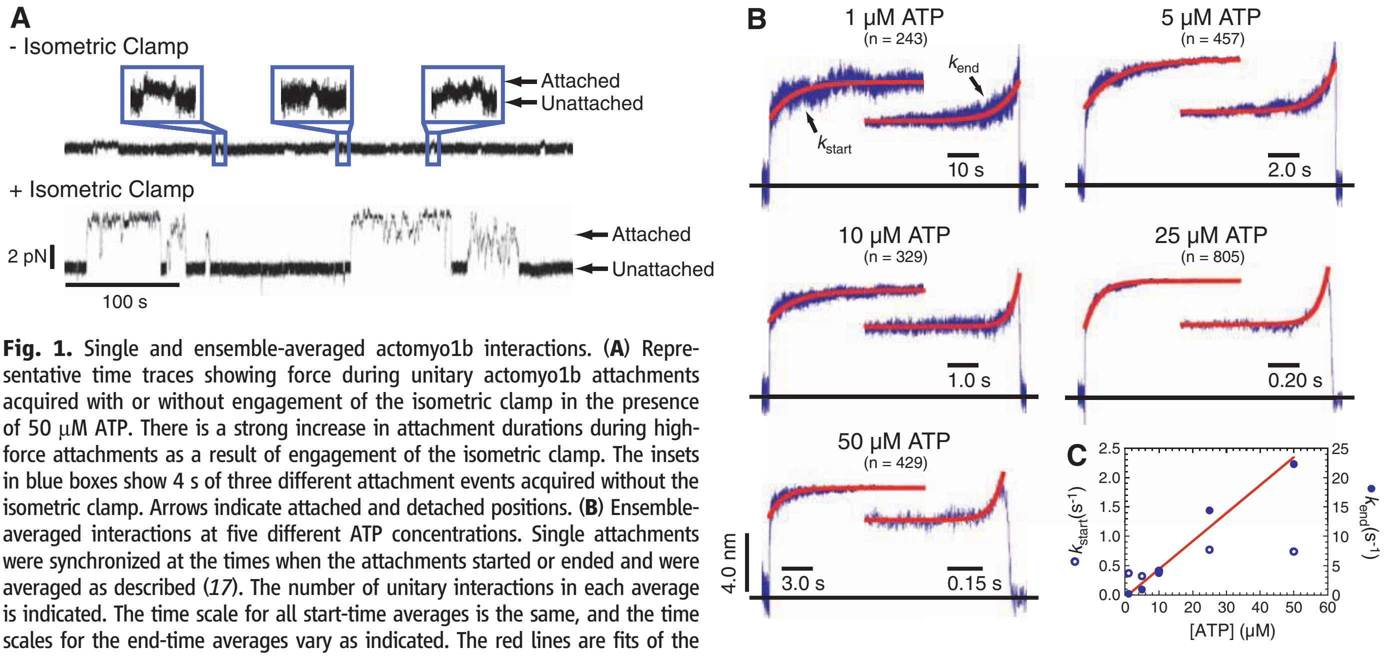


Fig. 1. Single and ensemble-averaged actomyo1b interactions. (**A**) Representative time traces showing force during unitary actomyo1b attachments acquired with or without engagement of the isometric clamp in the presence of 50 μM ATP. There is a strong increase in attachment durations during high-force attachments as a result of engagement of the isometric clamp. The insets in blue boxes show 4 s of three different attachment events acquired without the isometric clamp. Arrows indicate attached and detached positions. (**B**) Ensemble-averaged interactions at five different ATP concentrations. Single attachments were synchronized at the times when the attachments started or ended and were averaged as described (*17*). The number of unitary interactions in each average is indicated. The time scale for all start-time averages is the same, and the time scales for the end-time averages vary as indicated. The red lines are fits of the start-time (k_{start}) and end-time (k_{end}) averages to a single exponential rate function. (**C**) ATP concentration dependence of k_{start} (open circles) and k_{end} (solid circles) rates. The red line is a linear fit of the k_{end} rates, yielding a slope of 0.48 μM^{-1} s^{-1} (correlation coefficient = 0.985).

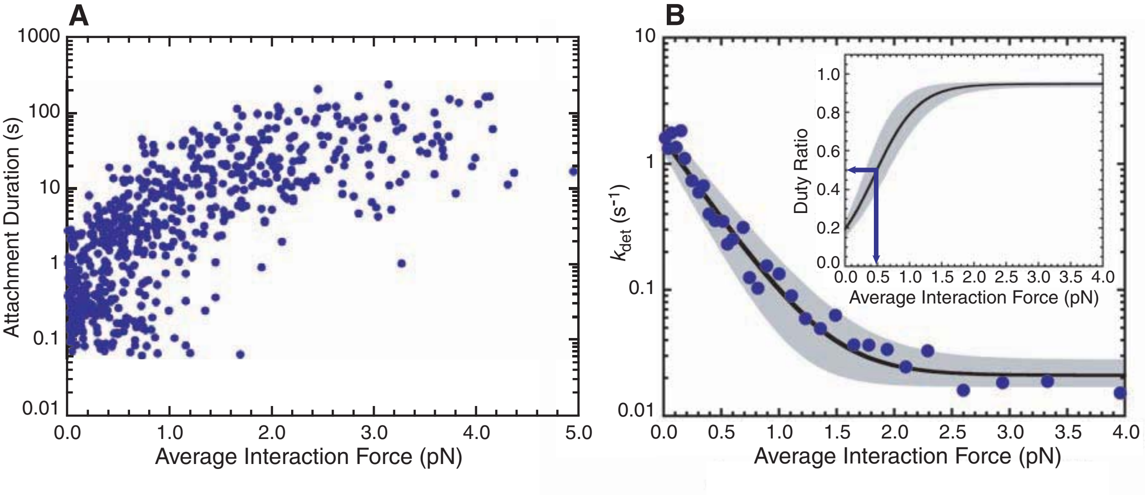


Fig. 2. Force dependence of actomyo1b attachment durations. (**A**) Attachment durations (n = 638 from 12 experiments) plotted as a function of the average force during the attachment in the presence of 50 μM ATP. (**B**) Force dependence of the actin-detachment rate (k_{det}) as calculated using the parameters from the fit of the data in (A) to Eq. 2 as determined by MLE. The best-fit parameters are k_{g0} = 1.6 s^{-1} (+0.5/–0.35 s^{-1}), d_{det} = 12 nm (+1.6/–3.0 nm), and k_i = 0.021 s^{-1} (+0.007/–0.004 s^{-1}). Points are the inverse averages of 20 consecutive attachment durations from the plot in (A). (Inset) Predicted myo1b duty ratio as a function of force. The shaded areas show the 97% confidence limit based on the uncertainties listed above. Arrows show the force at which myo1b becomes a high–duty-ratio motor (~0.5 pN).

stroke is the sum of two substeps, we examined the substep sizes and kinetics of the actomyosin interactions by ensemble-averaging the time courses of individual actomyo1b interactions that were synchronized at the times when the interactions started or ended (Fig. 1B). The time courses of the start-time averages reveal the lifetimes of the first substep, and the time courses of the end-time averages reveal the lifetimes of the second substep (*18*).

The time courses of the start-time averages have rapid initial 5.1 ± 0.43–nm substeps of actin displacement followed by slower 3.3 ± 0.35–nm increases to the final displacement. If the 3.3-nm substep is the result of a 32° lever-arm rotation (*11*), we calculate the effective myo1b lever-arm length to be 6.0 ± 0.63 nm. This length is shorter than expected given that the regulatory domain (the lever-arm region) contains five calmodulin-binding IQ motifs, which would have a length of ~20 nm if the lever arm were rigid (*19*). The short effective lever-arm of myo1b is probably due to weak calmodulin binding to a subset of the IQ motifs, resulting in a flexible regulatory domain (*16*). The regulatory domain is alternatively spliced in vivo (*20*), so it is possible that the compliance of this region is transcriptionally regulated so as to modulate the mechanical properties of this motor (*16*).

The initial 5.1-nm working-stroke substep occurred within the experimental response time and most likely corresponds to the displacement that accompanies phosphate (Pi) release (*10*). The time courses of the 3.3-nm increase in the start-time averages reveal the lifetimes of the 5.1-nm substep at the different ATP concentrations. These time courses, at all ATP concentrations, were well fit by a single exponential function with rates (k_{start} = 0.37 to 0.77 s^{-1}) slower than reported for ADP release from actomyo1b in solution [1.8 s^{-1} (*13*)]. Small loads on the actomyo1b interactions probably lead to these slower rates.

The time courses of end-time averages have the same initial and total displacements observed in the start-time averages (Fig. 1B), and they reveal the lifetimes of attachment after the 3.3-nm substep. The time courses were well fit by single exponential functions with rates (k_{end}) that are linearly related to the ATP concentration with a slope of 0.48 $\mu\mathrm{M}^{-1}\ \mathrm{s}^{-1}$ (Fig. 1C), which is similar to solution measurements of the apparent second-order rate constant for ATP binding [0.22 $\mu\mathrm{M}^{-1}\ \mathrm{s}^{-1}$ (*13*)]. A two-step model for myo1b detachment from actin with the rates obtained from the ensemble-averaged interactions describes the distributions of actomyo1b attachment lifetimes (fig. S2). These data confirm that the myo1b working stroke occurs in two substeps (*10*) with lifetimes consistent with solution kinetic measurements (*13*) of the rates of ADP release (lifetime of the 5.1-nm substep) and ATP binding (lifetime of the 3.3-nm substep).

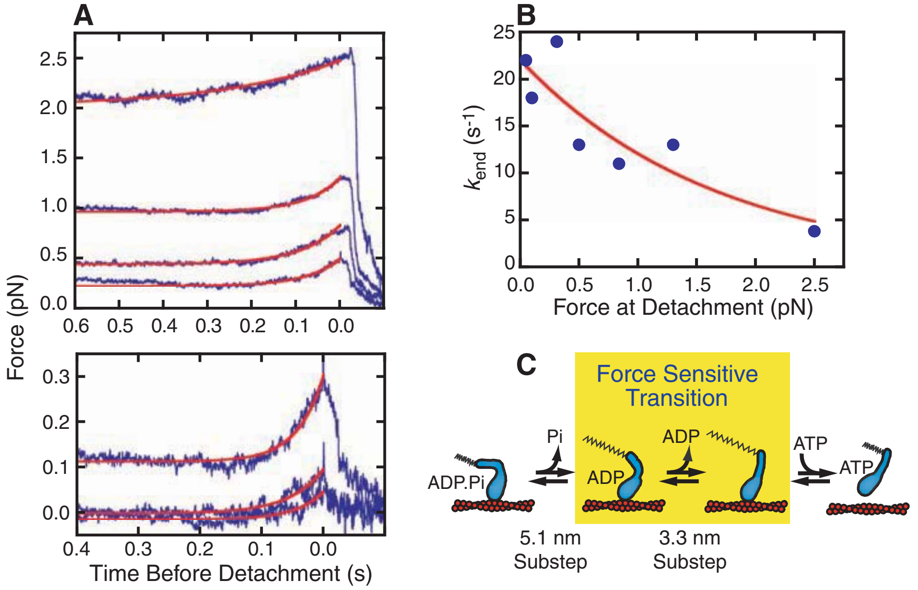


Fig. 3. Effect of force on working-stroke substep lifetimes. (**A**) Single interactions acquired with the isometric clamp in the presence of 50 μM ATP were sorted into bins based on the force immediately before detachment, synchronized to the interaction end-time, and averaged. From the bottom, the force bins (in piconewtons) are (0 to 0.125), (0.125 to 0.25), (0.25 to 0.50), (0.50 to 0.75), (0.74 to 1.0), (1.0 to 2.0), and (2.0 to 4.0). For clarity, the interaction averages in the three lowest force bins are shown on an expanded scale. The red lines are fits of the end-time averages to a single exponential rate function (k_{end}). The rates of the end-time averages are faster than the feedback response time of the isometric clamp, resulting in the lower force amplitudes for the faster substep time courses. (**B**) Force dependence of k_{end} rates obtained from the fits in (A). The red line is the best fit of the data to Eq. 4 with $k_{\mathrm{end0}} = 22 \pm 2.5\ \mathrm{s}^{-1}$ and $d_{\mathrm{end}} = 2.5 \pm 0.83$ nm. (**C**) Model for myo1b (blue) bound to actin (red) undergoing a working stroke. The 5.1-nm substep, 3.3-nm substep, and force-sensitive transition are identified. The rate of the ADP release step as a function of force is defined as k_{det} (Eq. 2), and the rate of ATP binding and subsequent actomyo1b detachment is defined as k_{end} (Eq. 4). The extended spring signifies tension on the actomyo1b complex.

The force dependence of actomyo1b attachment lifetimes was measured with a feedback system that applies a dynamic load to the actomyo1b to keep the actin near its isometric position while myosin undergoes its working stroke (*21*). With the feedback system, the resisting force applied to the actomyo1b attachment was determined primarily by the size of the myo1b powerstroke, the stiffness of the myo1b lever arm, and the position at which myosin binds to the actin filament.

We observed dramatic increases in actomyo1b attachment durations in the presence of 50 μM ATP upon engagement of the isometric clamp, with many durations exceeding 1 min (Fig. 2A). The attachment durations increased with increasing force until ~1.5 pN, after which they appeared to be force-independent. We assumed a model for the rate of actomyo1b detachment that includes force-dependent and force-independent pathways

$$\begin{tikzcd} \text{Attached} \arrow[r, "k_g"] \arrow[d, "k_i"'] & \text{Detached} \\ \text{Detached} & \end{tikzcd} \qquad (1)$$

where k_g is a force-dependent rate constant and k_i is a force-independent rate constant for actomyo1b dissociation. The force dependence of the detachment rate (k_{det}) can be calculated as (*22*)

$$k_{\mathrm{det}}(F) = k_{g0} e^{-\frac{F \cdot d_{\mathrm{det}}}{kT}} + k_i \qquad (2)$$

where k_{g0} is the rate of k_g in the absence of force, d_{det} is the distance parameter (the distance to the transition state), F is force, k is the Boltzmann constant, and T is the temperature. Because the attachment durations at each force are expected to be exponentially distributed, we used bootstrap Monte Carlo simulations to simulate data to use in maximum likelihood estimations (MLEs) (*17*). From the MLEs, we determined the values and confidence limits of d_{det}, k_{g0}, and k_i that best describe the distribution of attachment durations. The best-fit value of $k_{g0} = 1.6\ \mathrm{s}^{-1}$ (+0.5/–0.35 s^{-1}) is consistent with the rate of ADP release measured via biochemical methods [1.8 s^{-1} (*13*)], which limits k_{det} at 50 μM ATP in the absence of force. The distance parameter d_{det} = 12 nm (+1.6/–3.0 nm), which is a measure of strain sensitivity, is very large and distinguishes myo1b as an extraordinarily strain-dependent motor at loads <2 pN. This sensitivity is very different from that of other characterized myosins, including strain-sensitive myosin VI (*23*). Unlike myo1b, the strain sensitivity of myosin VI is seen only at low forces (<2 pN) in the presence of ADP.

The strain sensitivity of myo1b is most clearly illustrated in a plot of k_{det} versus force (Fig. 2B), where it can be seen that k_{det} decreases >75-fold with <2 pN of resisting force. At the low forces experienced by myo1b in the absence of the isometric clamp (<0.2 pN, Fig. 1B), a 1.5-fold decrease in k_{det} is predicted over the unloaded rate (k_{g0}), which is consistent with the values of k_{start} measured from the start-time averages.

The fraction of the total ATPase cycle in which myo1b is bound to actin in a force-bearing state is termed the duty ratio, and the force dependence of the duty ratio can be calculated as

$$\text{duty ratio}(F) = \frac{k_{att}}{k_{att} + k_{det}(F)} \qquad (3)$$

where k_{att} is the rate of entry into the strong binding states. k_{att} cannot be determined directly from the force time courses but can be estimated from the rate of phosphate release ($k_{att} = 0.38\ s^{-1}$, fig. S3). Myo1b transforms from a low–duty-ratio motor (<0.2) to a high–duty-ratio motor (>0.5) when working against as little as 0.5 pN of force, and it approaches the duty ratio of processive myosins (>0.9) at forces as low as 1.5 pN (Fig. 2B, inset).

We investigated the effect of force on the lifetimes and force amplitudes of the working-stroke substeps. Interactions acquired in the presence of 50 μM ATP were sorted into bins based on the force immediately before detachment, and individual interactions were synchronized at their end times and ensemble-averaged. Transient increases in force due to substeps were observed in the 500 ms before detachment in all force bins (Fig. 3A). Single exponential fits of the time courses yielded rates that decreased with increasing force (Fig. 3B). The force dependence of the rates was fit to the equation

$$k_{end}(F) = k_{end0}e^{-\frac{F \cdot d_{end}}{kT}} \qquad (4)$$

where k_{end0} is the rate of the time course in the absence of force and d_{end} is the distance parameter for the substep (Fig. 3B). The best-fit rate of k_{end0} (22 ± 2.5 s^{-1}) is consistent with the rate of 50 μM ATP binding in the absence of resisting loads ($k_{end} = 24\ s^{-1}$, Fig. 1C), and the value $d_{end} = 2.5 \pm 0.83$ nm is much smaller than d_{det} (Fig. 2B). Therefore, the ATP binding step is not the force-dependent step that limits the rate of actomyo1b detachment. ADP release is the most likely candidate for the force-dependent transition (Fig. 3C). d_{det} is substantially larger than the size of the substep that correlates with ADP release (3.3 ± 0.35 nm, Fig. 1B), indicating that the force-sensitive transition state is not on a coordinate that is in line with a rigid lever arm rotation (*24*).

The presence of a substep in the end-time averages in all force bins indicates that actomyo1b detachment did not occur before ADP release, even at forces where the detachment rate is force-insensitive (> 1.5 pN; Fig. 2). The force-independent detachment rate (k_i in Eq. 1) is probably the result of accelerated detachment due to force fluctuations in the system. Decreases in force before the ADP-release substep are observed in the force-binned start- and end-time averages (fig. S4). Thus, when force transiently drops, there is an exponentially higher probability of ADP release (Fig. 2B), which is followed by rapid ATP binding and detachment.

Our results show that myosin I responds to small resisting loads (<2 pN) by dramatically increasing the actin-attachment lifetime more than 75-fold. This impressive tension sensitivity supports models that identify myosin I as the adaptation motor in mechanosensory hair cells (*2*, *3*). More generally, the load-dependent kinetics support a model in which myosin I's function to generate and sustain tension for extended time periods, rather than to rapidly transport cargos (fig. S5). This new understanding of myosin I mechanics allows a more rigorous assignment of this motor's molecular roles in controlling organelle morphology (*8*) and dynamics (*5*) in the wide variety of cells types in which it is expressed.

References and Notes

1. J. S. Berg, B. C. Powell, R. E. Cheney, *Mol. Biol. Cell* **12**, 780 (2001).
2. J. R. Holt *et al.*, *Cell* **108**, 371 (2002).
3. C. Batters, M. I. Wallace, L. M. Coluccio, J. E. Molloy, *Philos. Trans. R. Soc. London Ser. B* **359**, 1895 (2004).
4. P. G. Gillespie, J. L. Cyr, *Annu. Rev. Physiol.* **66**, 521 (2004).
5. A. Bose *et al.*, *Nature* **420**, 821 (2002).
6. R. E. McConnell, M. J. Tyska, *J. Cell Biol.* **177**, 671 (2007).
7. D. E. Hokanson, J. M. Laakso, T. Lin, D. Sept, E. M. Ostap, *Mol. Biol. Cell* **17**, 4856 (2006).
8. L. Salas-Cortes *et al.*, *J. Cell Sci.* **118**, 4823 (2005).
9. M. A. Geeves, C. Perreault-Micale, L. M. Coluccio, *J. Biol. Chem.* **275**, 21624 (2000).
10. C. Veigel *et al.*, *Nature* **398**, 530 (1999).
11. J. D. Jontes, E. M. Wilson-Kubalek, R. A. Milligan, *Nature* **378**, 751 (1995).
12. R. F. Siemankowski, M. O. Wiseman, H. D. White, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 658 (1985).
13. J. H. Lewis, T. Lin, D. E. Hokanson, E. M. Ostap, *Biochemistry* **45**, 11589 (2006).
14. C. Veigel, S. Schmitz, F. Wang, J. R. Sellers, *Nat. Cell Biol.* **7**, 861 (2005).
15. J. T. Finer, R. M. Simmons, J. A. Spudich, *Nature* **368**, 113 (1994).
16. T. Lin, N. Tang, E. M. Ostap, *J. Biol. Chem.* **280**, 41562 (2005).
17. See methods in supporting information on *Science* Online.
18. C. Veigel, J. E. Molloy, S. Schmitz, J. Kendrick-Jones, *Nat. Cell Biol.* **5**, 980 (2003).
19. D. M. Warshaw *et al.*, *J. Biol. Chem.* **275**, 37167 (2000).
20. C. Ruppert, R. Kroschewski, M. Bahler, *J. Cell Biol.* **120**, 1393 (1993).
21. Y. Takagi, E. E. Homsher, Y. E. Goldman, H. Shuman, *Biophys. J.* **90**, 1295 (2006).
22. G. I. Bell, *Science* **200**, 618 (1978).
23. D. Altman, H. L. Sweeney, J. A. Spudich, *Cell* **116**, 737 (2004).
24. D. Tsygankov, M. E. Fisher, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19321 (2007).

25. We thank T. Lin for technical assistance and Y. E. Goldman for helpful discussions. Supported by NIH grants AR051174 (H.S. and E.M.O.) and GM057247 (E.M.O.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/133/DC1
Materials and Methods
Figs. S1 to S5
References

21 April 2008; accepted 9 June 2008
10.1126/science.1159419

The Spread of Ras Activity Triggered by Activation of a Single Dendritic Spine

Christopher D. Harvey,[1,2]* Ryohei Yasuda,[2,3]*† Haining Zhong,[1,2] Karel Svoboda[1,2]†

In neurons, individual dendritic spines isolate *N*-methyl-D-aspartate (NMDA) receptor–mediated calcium ion (Ca^{2+}) accumulations from the dendrite and other spines. However, the extent to which spines compartmentalize signaling events downstream of Ca^{2+} influx is not known. We combined two-photon fluorescence lifetime imaging with two-photon glutamate uncaging to image the activity of the small guanosine triphosphatase Ras after NMDA receptor activation at individual spines. Induction of long-term potentiation (LTP) triggered robust Ca^{2+}-dependent Ras activation in single spines that decayed in ~5 minutes. Ras activity spread over ~10 micrometers of dendrite and invaded neighboring spines by diffusion. The spread of Ras-dependent signaling was necessary for the local regulation of the threshold for LTP induction. Thus, Ca^{2+}-dependent synaptic signals can spread to couple multiple synapses on short stretches of dendrite.

Dendritic spines, small (<1 μm^3) protrusions emanating from the dendritic shaft, are the sites of most excitatory synapses in the mammalian brain (*1*). Spines function as biochemical compartments (*2*, *3*) that isolate postsynaptic Ca^{2+} accumulations (*4*–*6*). Ca^{2+} influx through synaptic *N*-methyl-D-aspartate (NMDA) receptors (NMDA-Rs) activates a complex signaling network (*7*), including the small guanosine triphosphatase (GTPase) H-Ras (*8*–*10*), to induce long-term potentiation (LTP) of synaptic transmission (*11*, *12*). LTP is input specific (*13*), suggesting that important Ca^{2+}-dependent signals remain confined to single spines. In contrast, synapses interact through diffusible cytoplasmic factors (*14*–*16*). Which signals downstream of NMDA-R–dependent Ca^{2+} influx are restricted to individual spines? To begin to address this question, we imaged the dynamics of Ras activity during the induction of input-specific LTP.

We transfected pyramidal neurons in organotypic hippocampal slices with a fluorescence resonance energy transfer (FRET)–based indicator of Ras activation, FRas-F, consisting of H-Ras tagged with monomeric enhanced green fluorescent protein (mEGFP) and the Ras-binding domain [RBD, R59A (mutation of Arg^{59} to Ala)] of Raf tagged with two monomeric red fluorescent proteins (mRFPs) (Fig. 1A) (*17*). Upon Ras activation, the affinity between Ras and RBD increases, leading to FRET between the donor and acceptor fluorophores (Fig. 1A) (*17*–*19*). FRas-F is rapidly reversible and reports the time course of endogenous Ras activation (*17*). We imaged FRET by using two-photon fluorescence lifetime imaging (2pFLIM) (*17*, *20*, *21*). To quantify Ras activation, we computed the fraction of Ras molecules binding to RBD (binding fraction) (Fig. 1B) (*17*, *22*).

To induce synapse-specific plasticity, we applied a train of two-photon glutamate uncaging pulses (30 pulses at 0.5 Hz) to a single spine in a low concentration (nominally 0 mM) of extracellular Mg^{2+} (*13*, *16*). Each uncaging pulse produced transient changes in the concentration of calcium ($[Ca^{2+}]$) and NMDA-R–mediated currents (7.3 ± 0.6 pA, corresponding to the opening of about five NMDA-Rs), similar to those triggered by low-frequency synaptic stimulation (fig. S1) (*6*, *22*, *23*). Uncaging-evoked $[Ca^{2+}]$ accumulations were restricted mostly to the heads of the stimulated spines (fig. S1, A to E) (*22*). The uncaging train caused a sustained spine enlargement in the stimulated spine; neighboring spines less than 4 μm away did not change (Fig. 1, C to E) (*13*, *16*). The increase in spine volume was proportional to an enhancement in postsynaptic sensitivity to glutamate, indicating that spine enlargement is a structural correlate of LTP (fig. S2C) (*13*, *16*, *24*).

The uncaging train induced robust Ras activation in the stimulated spine (Fig. 1, C, F, and G), which peaked within 1 min after the stimulus and returned to baseline levels within 15 min (decay time constant of Ras activation, τ_{decay} = 5.6 ± 0.5 min). Ras activation required Ca^{2+} influx through NMDA-Rs (Fig. 1H) (*8*–*10*). Ras activation also required the activity of multiple signaling factors; inhibitors of calcium/calmodulin-dependent protein kinase II (CaMKII; 10 μM KN62), phosphoinositide 3-kinase (PI3K; 20 μM LY294002), or protein kinase C (PKC; 1 μM Gö6976) signaling reduced Ras activation (Fig. 1H and fig. S3D) (*11*, *22*, *25*). The amplitudes of Ras activation and sustained spine enlargement were correlated (fig. S4A). Expression of a dominant-negative form of FRas-F [Ras S17N (mutation of Ser^{17} to Asn)] or inhibition of extracellular signal–regulated kinase (ERK) activation [mitogen-activated protein kinase (MAPK) kinase (MEK) blocker; 10 μM U0126] reduced the magnitude of sustained spine enlargement (fig. S4F) (*22*). Ras-ERK activation, therefore, was necessary for the persistent increase in spine volume, confirming the role of Ras signaling in synaptic plasticity (*11*). CaMKII activity, PKC signaling, and actin polymerization were also required for spine structural plasticity (fig. S4F) (*13*, *22*).

We next characterized the spatial profile of Ras activation (Figs. 1, F and G, and 2, A and B). After the plasticity-inducing stimulus, Ras activity spread over several micrometers in both directions along the parent dendrite and invaded nearby spines (length constant, $L \approx 11$ μm at 4 min) (Fig. 2B), suggesting that Ras signaling is not synapse specific.

Could the presence of the Ras sensor distort the spatial profile of Ras activity? The mean distance active Ras travels before it is inactivated, L, depends on the effective diffusion coefficient of Ras, D, and the time constant of Ras inactivation, $\tau_{inactivation}$ (*22*):

$$L \sim \sqrt{D\tau_{inactivation}} \qquad (1)$$

FRas-F expression could increase D and $\tau_{inactivation}$ by saturating Ras scaffolds and Ras inactivators

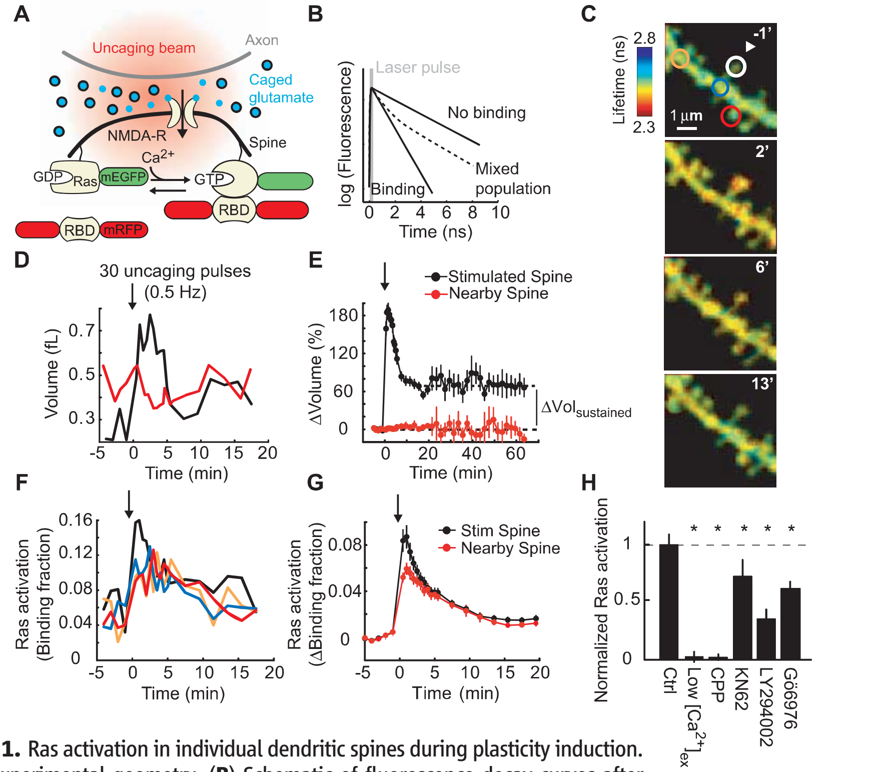


Fig. 1. Ras activation in individual dendritic spines during plasticity induction. (**A**) Experimental geometry. (**B**) Schematic of fluorescence decay curves after pulsed excitation. Slow and fast components correspond to free donor and donor bound to acceptor, respectively. FRET decreases fluorescence lifetime. (**C**) Fluorescence lifetime images of Ras activity. At time = 0, 30 uncaging pulses (0.5 Hz) were applied to the spine marked by the arrowhead in a low concentration (nominally 0 mM) of extracellular Mg^{2+}. "Warmer" colors indicate shorter lifetimes and higher levels of Ras activation. (**D**) Changes in spine volume. Colors correspond to the circles in (C). Arrow, time of stimulus. (**E**) Spine volume changes for the stimulated and nearby (<4 μm) spines (–5 to 20 min: 91 spines; >20 min: 9 spines, mean ± SEM). $\Delta Vol_{sustained}$ is the volume difference between 15.5 and 19.5 min and the baseline volume. (**F**) Ras activation. Colors correspond to the circles in (C). (**G**) Ras activation in the stimulated and nearby spines (82 spines, mean ± SEM). (**H**) Pathways to Ras activation. Ras activation was the average binding fraction at 1 to 3 min minus baseline, normalized to the control condition. Numbers of spines: 82 Ctrl, 11 Low $[Ca^{2+}]_{ex}$ (200 μM), 12 CPP (10 μM), 27 KN62 (10 μM), 20 LY294002 (20 μM), and 23 Gö6976 (1 μM). Error bars indicate mean ± SEM. *$P < 0.05$ versus control.

[1] Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147, USA. [2] Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [3] Neurobiology Department, Duke University Medical Center, Durham, NC 27710, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: yasuda@neuro.duke.edu, svobodak@janelia.hhmi.org

(GTPase-activating proteins, GAPs), respectively. In addition, RBD competes with GAPs for binding to Ras, further increasing $\tau_{\text{inactivation}}$ (*17*). To determine if FRas-F expression increases the spread of Ras activation (fig. S5) (*22*), we modulated FRas-F levels by changing the duration of expression and imaged Ras activation. We quantified FRas-F expression by using post hoc immunofluorescence measurements (fig. S6) (*22*). Ras activation spread along the dendrite even at the lowest FRas-F expression levels tested (~1.5-fold Ras overexpression) (Fig. 2, C and D). Furthermore, the spatial profile of Ras activity was similar across a wide range of FRas-F concentrations (Fig. 2, C and D). FRas-F expression, therefore, did not distort the spatial spread of Ras activation.

We next investigated the mechanisms underlying the spread of Ras activity. Ras mobility could be a major determinant of the spread of Ras activity (Eq. 1). To measure mobility directly, we expressed H-Ras tagged with photoactivatable GFP (paGFP-Ras) (*26*) together with the cytoplasm marker mCherry. After photoactivation in single spines, the spine fluorescence decayed over seconds ($\tau = 4.17 \pm 0.40$ s), similar to the decay rate of membrane-targeted paGFP and less than one-tenth the decay rate of cytosolic paGFP (Fig. 3, A and B) (*3*). These measurements suggest that Ras diffuses relatively freely within the plasma membrane. Constitutively active Ras [Ras G12V (mutation of Gly^{12} to Val)] diffused at a rate similar to that of wild-type Ras (Fig. 3B) and had similar mobility with or without RBD expression (*22*); Ras activation and the binding of RBD to Ras, therefore, did not alter Ras diffusion appreciably. Near physiological temperatures (33°C), Ras mobility was enhanced by a factor of ~2, likely as a result of changes in membrane viscosity (Fig. 3B) (*27*). Within 30 s after photoactivation, more than 90% of the paGFP-Ras G12V fluorescence had decayed from the spine head (Fig. 3D). A similar fraction of membrane-targeted paGFP remained in the spine after 30 s (Fig. 3D). These studies imply that no major immobile fraction of active Ras exists in the spine. Ras overexpression could enhance diffusion by saturating Ras scaffolds (fig. S5A) (*22*, *28*). However, the decay time constant and the fraction of fluorescence remaining in the spine at 30 s were independent of paGFP-Ras G12V expression levels (Fig. 3, C and E). These data indicate that active endogenous Ras is highly mobile and does not associate strongly with immobile scaffolds in the spine.

The time constant of Ras inactivation could also modulate the spread of Ras activity (Eq. 1). Because the decay of Ras activity following uncaging stimuli includes both Ras inactivation and diffusion, we instead measured Ras inactivation after brief trains of back-propagating action potentials (bAPs), which activate Ras globally (Fig. 3, F and G) (*17*). Because the length constant of Ras diffusion (~10 µm) (Fig. 2B) is much longer than interspine distances, Ras molecules

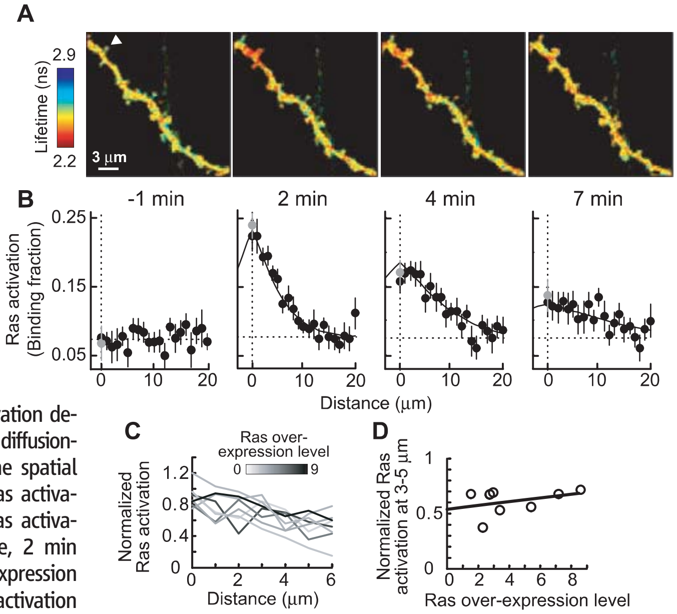


Fig. 2. Spatial spread of Ras activity. (**A**) Fluorescence lifetime images of Ras activity. At time = 0, LTP was induced at the spine marked by an arrowhead. (**B**) The spatial spread of Ras activation at different time points. Black circles indicate distances along the dendrite relative to the stimulated spine (gray circle; $n = 11$ cells, mean ± SEM). The solid line shows the fitted profile of Ras activation derived from a one-dimensional diffusion-reaction model (*22*). (**C**) The spatial spread of Ras activation (Ras activation, normalized to peak Ras activation in the stimulated spine, 2 min after stimulus) at various expression levels. (**D**) Normalized Ras activation 3 to 5 µm from the stimulated spine. $r = 0.36$, $P > 0.3$, $n = 8$ cells.

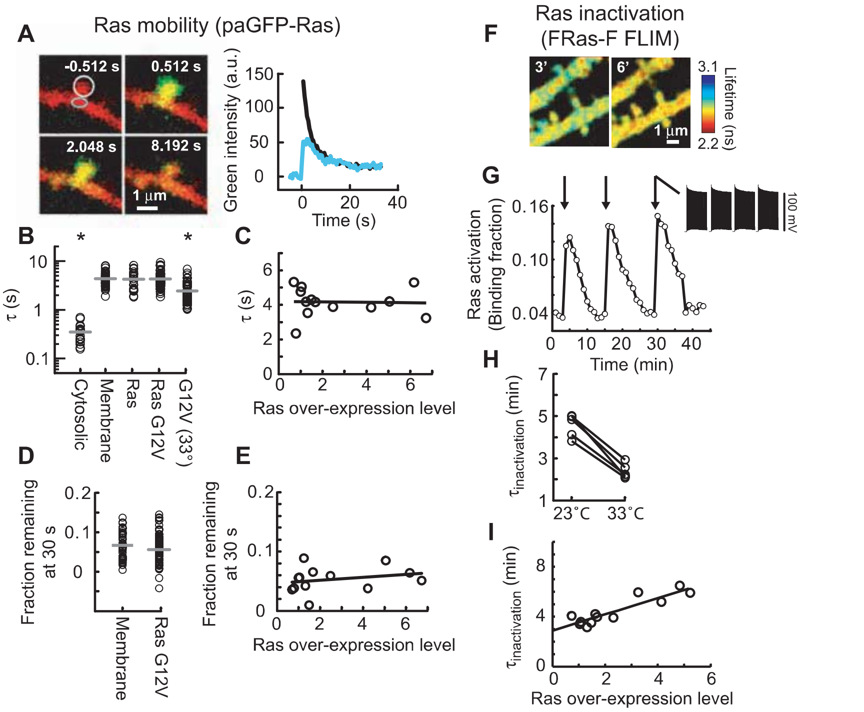


Fig. 3. Ras mobility and the time constant of Ras inactivation. (**A**) Left: Spine before and after photoactivation (green, paGFP-Ras; red, mCherry). Right: Time course of activated paGFP-Ras in the spine (black) and parent dendrite (blue). (**B**) Decay time constants of paGFP fluorescence in the photoactivated spine. Horizontal bars, mean. Numbers of spines: 17 cytosolic, 46 membrane, 21 Ras, 84 Ras G12V, 52 Ras G12V (33°C). *$P < 0.05$ versus paGFP-Ras G12V (23°C). (**C**) Decay time constants of paGFP-Ras G12V fluorescence at varying expression levels. $r = -0.05$, $P > 0.8$, $n = 13$ cells. (**D**) Fraction of paGFP fluorescence remaining in the spine 30 s after photoactivation. Horizontal bars, mean. Number of spines: 46 membrane, 84 Ras G12V. $P > 0.7$. (**E**) Fraction of paGFP-Ras G12V fluorescence remaining in the spine at 30 s at varying expression levels. $r = 0.25$, $P > 0.4$, $n = 13$ cells. (**F**) Fluorescence lifetime images before and after trains of bAPs (40 APs at 83 Hz, repeated four times every 5 s). (**G**) Time course of Ras activation for the experiment shown in (F). Arrows, AP stimuli. (**H**) $\tau_{\text{inactivation}}$ at 23°C and 33°C. Ras inactivation was measured at both temperatures in individual cells. $n = 5$ cells, $P < 0.001$. (**I**) Ras inactivation at varying expression levels. $r = 0.8$, $P < 0.01$, slope = 0.6 min per fold overexpression, intercept = 2.9 min. $n = 12$ cells.

activated in both dendrites and spines sample the same population of GAPs, independent of the stimulus. After trains of bAPs, Ras activity decayed over minutes ($\tau_{\text{inactivation}} = 4.5 \pm 0.4$ min) (Fig. 3G) (*17*). Near physiological temperatures (33°C), $\tau_{\text{inactivation}}$ was decreased by a factor of ~2 (Fig. 3H). Because this reduction in $\tau_{\text{inactivation}}$ was similar to the enhancement of Ras mobility (Fig. 3B), the spread of Ras activity is likely to be relatively independent of temperature (Eq. 1). The duration of Ras activity may be prolonged by FRas-F expression (fig. S5, B and C) (*22*). However, $\tau_{\text{inactivation}}$ was only weakly dependent on FRas-F expression levels (Fig. 3I). Extrapolation to the native case gave $\tau_{\text{inactivation}} \sim 2.9$ min (Fig. 3I). Because RBD and exogenous Ras levels were correlated ($r = 0.89$, $P < 0.001$), the estimation of $\tau_{\text{inactivation}}$ takes into account the effects of both exogenous Ras and RBD expression. FRas-F therefore increased $\tau_{\text{inactivation}}$ by a factor of ~2 and thus the spatial spread by a factor of ~1.5 (Eq. 1). Active Ras in the native case is therefore expected to spread by diffusion over ~10 μm of dendrite, invading 10 to 20 synapses (*22*).

To address whether diffusion of active Ras by itself can quantitatively explain the spread of Ras activity, we modeled the spread of Ras activation using a one-dimensional diffusion-reaction model. In the model, Ras is activated by stationary guanine nucleotide exchange factors (GEFs) in the stimulated spine and is inactivated by GAPs that are distributed homogeneously along the dendrite. By fitting our spatial Ras activity data (Fig. 2B) using a global regression analysis, we obtained estimates of the time constant of GEF activity, $\tau_{\text{GEF}} = 2.0$ min, $\tau_{\text{inactivation}} = 4.5$ min, and $D = 0.65\ \mu\text{m}^2/\text{s}$. These estimates of $\tau_{\text{inactivation}}$ and D are identical to the values measured independently (Fig. 3) (*22*, *29*). We conclude that the spread of Ras activation can be explained by the rapid diffusion of active Ras.

What could be the function of the spread of Ras activation? Induction of LTP at a single synapse reduces the threshold for potentiation at neighboring synapses (*16*). Because the time scales (~5 min) and length scales (~10 μm) of Ras signaling and synaptic crosstalk (*16*) are identical, we tested whether the spread of Ras activity is critical for the local modulation of the LTP induction threshold.

In a synaptic crosstalk experiment, LTP is induced at one spine (LTP spine) with a train of uncaging pulses; subsequently, a subthreshold uncaging protocol, which by itself is too weak to trigger plasticity (fig. S7C), is sufficient to induce functional and structural plasticity in a neighboring spine (sub spine) (*16*). If the spread of Ras activity is necessary for crosstalk, Ras-dependent signaling is likely required at a later time in the sub spine as compared to the LTP spine. We induced LTP at a single spine and a short time (3 min) later locally applied a MEK inhibitor (20 μM U0126) to block Ras-ERK signaling; the inhibitor did not decrease the sustained spine-volume change in the LTP spine appreciably (Fig. 4, B to D). We then stimulated a neighboring spine with the subthreshold protocol. The MEK inhibitor substantially reduced the sustained spine enlargement in the sub spine (Fig. 4, B to D), indicating that Ras-dependent signaling is necessary for crosstalk in plasticity.

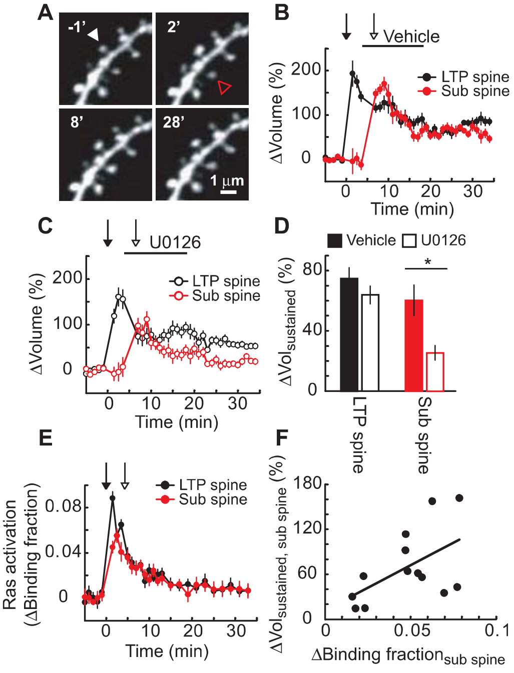


Fig. 4. Spread of Ras signaling and synaptic crosstalk. (**A**) Time-lapse images of a GFP-expressing pyramidal neuron in an acute hippocampal brain slice. At time = 0, 30 uncaging pulses (0.5 Hz, 4-ms pulse duration, LTP protocol) were applied to the spine marked by a white arrowhead (LTP spine) in a low concentration (nominally 0 mM) of extracellular Mg^{2+}. At time = 3 min, vehicle (0.1% dimethyl sulfoxide) was pressure applied locally from a glass pipette until time = 17 min. At time = 5.5 min, the subthreshold protocol (30 uncaging pulses, 0.5 Hz, 1-ms pulse duration) was applied to a nearby spine (sub spine, red arrowhead). (**B**) Spine-volume changes in the vehicle condition (11 spines, mean ± SEM). (**C**) Spine-volume changes in the 20 μM U0126 condition (11 spines, mean ± SEM). (**D**) Sustained changes in spine volume. Error bars indicate mean ± SEM. $^*P < 0.05$. (**E**) Ras activation in the LTP and sub spines during the crosstalk paradigm in cultured hippocampal slices. At time = 0, the LTP protocol was applied to the LTP spine, and 3.5 min later, the subthreshold protocol was applied to the sub spine (13 spines, mean ± SEM). (**F**) Relation between Ras activation and the sustained spine enlargement in the sub spine during crosstalk. ΔBinding fraction was measured before the subthreshold protocol (1.5 to 3.5 min after the LTP protocol). $r = 0.54$, $P = 0.05$.

Ras activity in the sub spine could be caused exclusively by the spread of Ras activity from the LTP spine, but may also include Ras directly activated in the sub spine by the subthreshold protocol. However, the subthreshold protocol alone did not induce detectable Ras activation (fig. S7B). Similarly, in the crosstalk stimulus paradigm, the subthreshold protocol did not trigger additional Ras activation in the sub spine; the decay of Ras activation triggered by the LTP protocol was similar with and without application of the subthreshold protocol (compare Figs. 1G and 4E) (*22*). Furthermore, the magnitude of Ras activation in the sub spine before the subthreshold protocol, due to the spread of Ras activity, was correlated with the sustained spine enlargement in the sub spine (Fig. 4F). Together, these data indicate that the spread of Ras-dependent signaling is necessary for the local regulation of the LTP induction threshold.

We have shown that Ras signaling after LTP induction spreads along 10 μm of dendrite and invades neighboring spines. The time scale of Ras inactivation was on the order of minutes, as compared with a spine-dendrite diffusional coupling time of seconds. Active Ras can therefore diffuse out of the spine head into the dendrite before inactivating. In contrast, Ca^{2+} is highly compartmentalized because it is extruded much more rapidly (tens of milliseconds) than the diffusional relaxation time between spine and dendrite (hundreds of milliseconds) (*6*). The diffusional barrier created by the narrow spine neck (*2*, *3*) may be important for rapid signals, such as Ca^{2+}, but is likely less significant for other diffusible signals with longer durations. Synapse-specific plasticity presumably is achieved by signaling molecules that are strongly tethered to postsynaptic scaffolds. The spread of Ca^{2+}-dependent signaling and local functional interactions at the level of plasticity (*16*) indicate that neighboring synapses are co-regulated. Synapses sharing a short stretch of dendrite may therefore form functional units within individual neurons (*30*).

References and Notes

1. E. A. Nimchinsky, B. L. Sabatini, K. Svoboda, *Annu. Rev. Physiol.* **64**, 313 (2002).
2. K. Svoboda, D. W. Tank, W. Denk, *Science* **272**, 716 (1996).
3. B. L. Bloodgood, B. L. Sabatini, *Science* **310**, 866 (2005).
4. W. Muller, J. A. Connor, *Nature* **354**, 73 (1991).
5. R. Yuste, W. Denk, *Nature* **375**, 682 (1995).
6. B. L. Sabatini, T. G. Oertner, K. Svoboda, *Neuron* **33**, 439 (2002).
7. M. B. Kennedy, H. C. Beale, H. J. Carlisle, L. R. Washburn, *Nat. Rev. Neurosci.* **6**, 423 (2005).
8. M. J. Kim, A. W. Dunah, Y. T. Wang, M. Sheng, *Neuron* **46**, 745 (2005).
9. X. Tian *et al.*, *EMBO J.* **23**, 1567 (2004).
10. H. Y. Yun, M. Gonzalez-Zulueta, V. L. Dawson, T. M. Dawson, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5773 (1998).
11. J. J. Zhu, Y. Qin, M. Zhao, L. Van Aelst, R. Malinow, *Cell* **110**, 443 (2002).
12. R. C. Malenka, M. F. Bear, *Neuron* **44**, 5 (2004).
13. M. Matsuzaki, N. Honkura, G. C. Ellis-Davies, H. Kasai, *Nature* **429**, 761 (2004).
14. S. Tsuriel *et al.*, *PLoS Biol.* **4**, e271 (2006).
15. N. W. Gray, R. M. Weimer, I. Bureau, K. Svoboda, *PLoS Biol.* **4**, e370 (2006).
16. C. D. Harvey, K. Svoboda, *Nature* **450**, 1195 (2007).

17. R. Yasuda *et al*., *Nat. Neurosci.* **9**, 283 (2006).
18. N. Mochizuki *et al*., *Nature* **411**, 1065 (2001).
19. O. Rocks *et al*., *Science* **307**, 1746 (2005).
20. D. W. Piston, D. R. Sandison, W. W. Webb, in *Time-Resolved Laser Spectroscopy in Biochemistry III*, J. R. Lakowicz, Ed. (SPIE, Bellingham, WA, 1992), pp. 379–389.
21. E. Gratton, S. Breusegem, J. Sutin, Q. Ruan, N. Barry, *J. Biomed. Opt.* **8**, 381 (2003).
22. Supporting online material is available on *Science* Online.
23. E. A. Nimchinsky, R. Yasuda, T. G. Oertner, K. Svoboda, *J. Neurosci.* **24**, 2054 (2004).
24. C. D. Kopec, B. Li, W. Wei, J. Boehm, R. Malinow, *J. Neurosci.* **26**, 2000 (2006).
25. M. Fivaz, S. Bandara, T. Inoue, T. Meyer, *Curr. Biol.* **18**, 44 (2008).
26. G. H. Patterson, J. Lippincott-Schwartz, *Science* **297**, 1873 (2002).
27. E. A. Reits, J. J. Neefjes, *Nat. Cell Biol.* **3**, E145 (2001).
28. H. Niv, O. Gutman, Y. Kloog, Y. I. Henis, *J. Cell Biol.* **157**, 865 (2002).
29. P. H. Lommerse *et al*., *Biophys. J.* **86**, 609 (2004).
30. A. Losonczy, J. K. Makara, J. C. Magee, *Nature* **452**, 436 (2008).
31. We thank N. Ghitani, B. Burbach, C. Zhang, B. Shields, and H. White for technical assistance; N. Gray for help with immunostaining; L. van Aelst for helpful discussions; and J. Dudman for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute, NIH, a David and Fanny Luke Fellowship (C.D.H.), Burroughs Wellcome Fund (R.Y.), Dana Foundation (R.Y.), National Alliance for Autism Research (R.Y.), National Institute of Mental Health (R.Y.), and National Alliance for Research on Schizophrenia and Depression (H.Z.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1159675/DC1
Materials and Methods
SOM Text
Figs. S1 to S12
References

28 April 2008; accepted 28 May 2008
Published online 12 June 2008;
10.1126/science.1159675
Include this information when citing this paper.

Finite Scale of Spatial Representation in the Hippocampus

Kirsten Brun Kjelstrup,[1] Trygve Solstad,[1] Vegard Heimly Brun,[1] Torkel Hafting,[1] Stefan Leutgeb,[1] Menno P. Witter,[1,2] Edvard I. Moser,[1]* May-Britt Moser[1]

To determine how spatial scale is represented in the pyramidal cell population of the hippocampus, we recorded neural activity at multiple longitudinal levels of this brain area while rats ran back and forth on an 18-meter-long linear track. CA3 cells had well-defined place fields at all levels. The scale of representation increased almost linearly from <1 meter at the dorsal pole to ~10 meters at the ventral pole. The results suggest that the place-cell map includes the entire hippocampus and that environments are represented in the hippocampus at a topographically graded but finite continuum of scales.

Although the basic intrinsic circuitry of the hippocampus is similar along the entire dorsoventral axis of the structure (*1*), dorsal and ventral regions may not have similar functions. Dorsal and intermediate regions are preferentially connected, via the dorsolateral and intermediate bands of the entorhinal cortex, to visual and somatosensory cortices important for accurate spatial navigation (*2*–*5*), and selective lesions in these hippocampal regions can lead to impairments in spatial learning (*6*–*8*). Pyramidal cells in these parts of the hippocampus have spatially selective firing fields that reflect the animal's location (*9*, *10*) and jointly form a maplike representation of the environment (*10*–*12*). The scale of this representation increases from dorsal to intermediate hippocampus (*13*, *14*), matching the progressive dorsoventral increase in the spacing of grid cells in medial entorhinal cortex, one synapse upstream (*15*, *16*). In contrast, the ventral one-third of the hippocampus has strong connections with a restricted number of interconnected areas of the lateral septum, hypothalamus, and amygdala, both directly and through the ventromedial portions of the entorhinal cortex (*2*, *17*–*19*). Selective lesions of the ventral hippocampus affect autonomic and defensive responses but not basic spatial behaviors (*20*, *21*). Collectively, these studies imply a functional division in the hippocampus where the dorsal and intermediate parts are more important for spatial behavior and the ventral part is more relevant for emotional and motivational processes. However, such a division does not rule out the possibility that location is represented at all levels of the hippocampus. Spatial inputs may reach even the most ventral cells via the associational networks of the hippocampus (*22*, *23*) or by way of intrinsic connections between dorsal and ventral parts of the entorhinal cortex (*3*). To determine whether the ventral hippocampus also has place cells, we compared neural activity at multiple longitudinal levels of CA3 in 21 rats (*24*). All recording sites were mapped onto the long axis of CA3 (fig. S1). The data were subdivided into three groups based on recording location (dorsal, 7 to 22%; intermediate, 40 to 70%; and ventral, 70 to 100%).

Conventional recording environments may be too small to visualize the most extended hippocampal representations (figs. S2 to S4 and supporting online text). Thus, we tested the animals on an 18-m-long linear track. Well-delineated place fields, defined as spatially stable contiguous regions with firing above 20% of the peak rate, could be found at all longitudinal levels of the hippocampus (Fig. 1 and fig. S5). The representations scaled up between dorsal and ventral recording locations (Fig. 1), from an average place field size of 0.98 ± 0.03 m in dorsal hippocampus (mean ± SEM; 111 cells, five rats), to 3.73 ± 0.43 m in intermediate hippocampus (37 cells, three rats), and 5.52 ± 0.54 m in ventral hippocampus (61 cells, four rats). Few fields were longer than 10 m. The relation between recording location and field size was strong (dorsal versus intermediate versus ventral: $F_{2,517} = 86.5$, $P < 0.001$; all pairwise comparisons significant at $P < 0.001$, Bonferroni test). Forty-five per cent of the variance was explained by a linear regression model [$r = 0.67$, $n = 111$, $P < 0.001$; (fig. S6)]. There was a similar increase in the width of the central peak of the spatial autocorrelation function, measured at 20% of the peak value (dorsal, 1.12 ± 0.03 m; intermediate, 3.07 ± 0.19 m; and ventral, 4.10 ± 0.18 m; $F_{2,327} = 179.8$, $P < 0.001$; all pairwise comparisons significant at $P < 0.001$). In this case, 53% of the variance was explained by a linear regression model [$r = 0.73$, 111, $P < 0.001$; (fig. S6; supporting online text)].

The size of individual hippocampal place fields can also be estimated by exploiting the fact that place cells exhibit theta phase precession (*25*). A place field can be defined as the area between the points on a track where the precession begins and terminates (*26*). Running on the 18-m track was associated with theta rhythmicity and phase precession at all dorsoventral recording levels in CA3 (81 place cells; Fig. 2 and fig. S7). As the rat crossed the firing field, the firing phase advanced gradually, up to 360°, over successive cycles of the theta rhythm. The strength of precession was estimated for each firing field by rotating the phase by position distribution in steps of 1° across the phase cycle and fitting a linear regression line for each rotation (*25*, *27*). The rotation that gave the largest explained variance was identified. The slope of the regression line at this rotation (phase change per unit of movement) decreased as the field size increased along the long axis of the hippocampus (Fig. 2). In the ventral CA3, the phase advance often occurred over distances as large as 10 m (Fig. 2, right). The mean regression slope changed from –102° ± 33° per meter in dorsal hippocampus ($t_{62} = 3.1$, $P < 0.005$) to –18° ± 9° per meter in intermediate and ventral hippocampus ($t_{37} = 2.09$, $P < 0.05$; group difference: $t_{99} = 2.3$, $P < 0.05$; intermediate and ventral cells were combined be-

[1] Kavli Institute for Systems Neuroscience and Centre for the Biology of Memory, Norwegian University of Science and Technology, 7489 Trondheim, Norway. [2] Research Institute Neurosciences, Department of Anatomy and Neurosciences, VU University Medical Center, 1007 MB Amsterdam, Netherlands.

*To whom correspondence should be addressed. E-mail: edvard.moser@ntnu.no

cause of low numbers). The 5- to 10-fold increase in the distance covered by one cycle of precession is consistent with the 5- to 10-fold increase in the width of the place fields. Precession over such long distances also implies that the active firing regions in the ventral hippocampus are part of one place field rather than multiple overlapping fields.

The somewhat irregular shape of the ventral hippocampal place fields makes it difficult to estimate field size accurately because the algorithm sometimes cleaves fields as a result of local rate variations (see examples in Fig. 1 and fig. S5) and because fields at the end of the track may be truncated. To reduce the impact of these variables, we combined data from multiple recordings in a population vector analysis (Fig. 3 and fig. S8). Rate maps were stacked separately for left and right running directions, and the length of the track was segmented into 5-cm bins, giving a total of 330 population vectors per running direction after exclusion of the turning points. The correlation between all pairs of population vectors was plotted in a color-coded matrix where the correlation of each population vector with itself is represented along the diagonal (*14*, *28*). The width of the band of high correlations along the diagonal provides an indication of the size of a typical place field, without the constraints caused by cell-specific rate and shape variations. The mean half-width (at r = 0.50) was 0.81 m for dorsal CA3, 3.71 m for intermediate CA3, and 6.66 m for ventral CA3 (Fig. 3). The mean width at r = 0.20 was 1.41 m (dorsal), 8.65 m (intermediate), and 13.59 m (ventral). As was true for the analyses for individual cells, the relation between position on the dorsoventral axis and width of the correlated band was approximately linear (Fig. 4).

This study has two main findings. First, place cells exist along the entire longitudinal axis of the hippocampus, even in the most ventral region where the connectivity with somatosensory and

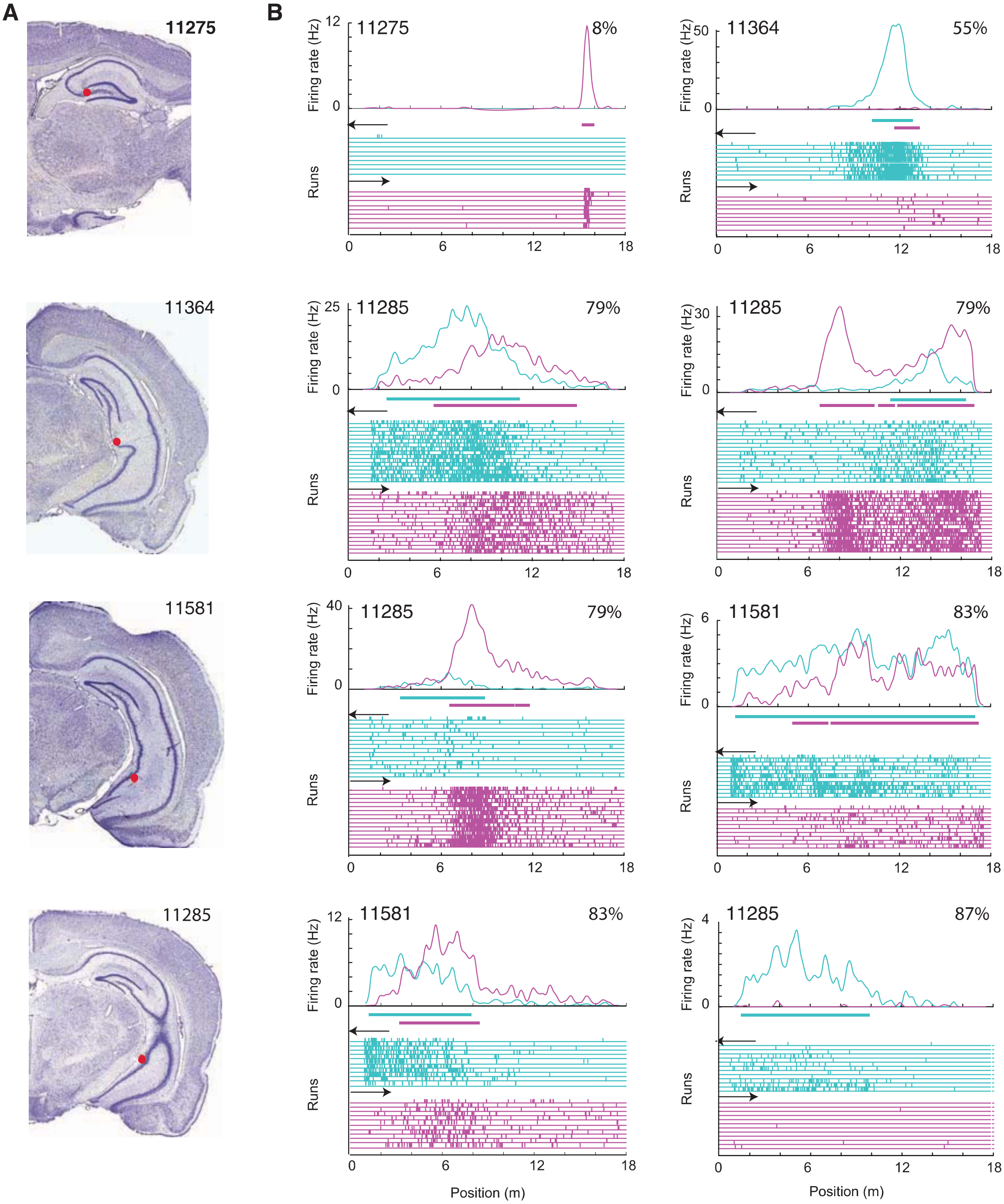


Fig. 1. Place fields of eight pyramidal cells recorded at different longitudinal levels of CA3 during animals' running on a linear 18-m track. (**A**) Nissl-stained sections showing recording locations in four animals. Individual rat numbers are indicated. (**B**) Place fields on the 18-m track. Each panel shows one cell. Rat numbers refer to (A). Percentages indicate location along the dorsoventral axis. For rat 11285, only the 87% location is shown in (A). (Top of each panel) Smoothed spike density function indicates firing rate as a function of position. Horizontal bar indicates estimated place field. Left runs, red; right runs, green. (Bottom of each panel) Raster plot showing density of spikes on individual laps. Each vertical tic indicates one spike and each horizontal line shows one lap (right side, blue-green; left, red). See fig. S5 for complete cell samples. Note 5- to 10-m-long place fields in ventral CA3 (colored horizontal bars; left runs, red; right runs, green).

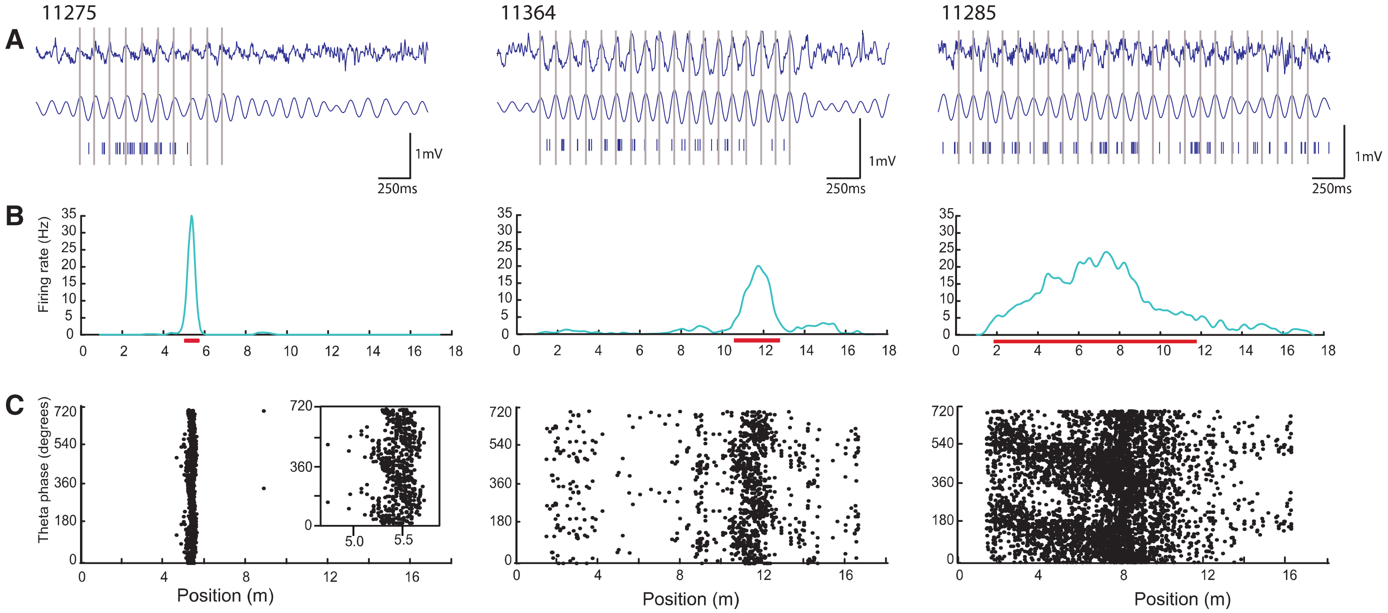


Fig. 2. Reduced rate of theta-phase precession in ventral hippocampus. Three cells are shown (left) dorsal CA3, (middle) intermediate CA3, and (right) ventral CA3. Rats and recording positions are the same as in Fig. 1. Only right runs are shown. (**A**) Local EEG with spike times (3.0 s). (Top trace) unfiltered; (bottom trace) 5 to 11 Hz filtered. Gray vertical lines indicate peak of theta cycle (0° or 360°). Individual spikes are shown as small vertical tics. (**B**) Smoothed spike density function showing size and location of firing field on the 18-m track. Red horizontal bar indicates estimated extent of place field. (**C**) Theta phase (two cycles) as a function of position. Each dot corresponds to one spike. For the dorsal cell, the phase-position field is magnified in the inset. Note 5- to 10-fold increase in the distance covered by one cycle of precession as field size increases toward the ventral hippocampus.

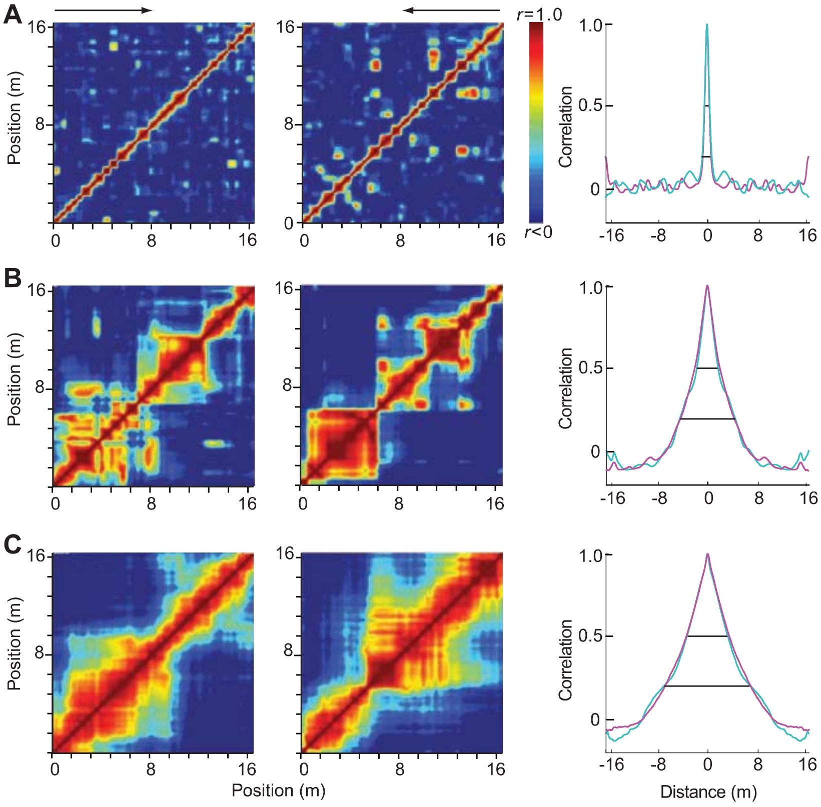


Fig. 3. Population vector autocorrelograms showing graded increase in spatial scale (**A**) dorsal, (**B**) intermediate, and (**C**) ventral CA3. (Left and middle) Linear rate maps of all pyramidal cells from each location were stacked, and a population vector was defined for each 5-cm bin of the composite map. Pearson's product moment correlation coefficient was calculated for each pair of population vectors, giving a two-dimensional correlation matrix. Correlation is color-coded (for *r* between –1 and 0, color is blue; color scale from 0 to 1 is linear, as shown on the scale bar). The width of the diagonal color band indicates the distance over which successive population vectors become decorrelated (*14*, *28*). Arrows indicate running direction; see fig. S8 for both directions. (Line graphs) Correlation as a function of spatial distance between population vectors. Horizontal lines indicate half-width (*14*) and 20% width.

visual brain regions is less direct (*2*–*5*, *29*). The origin of the spatial inputs to the ventral hippocampus is not known, but spatial information can probably reach ventral parts of the hippocampus both by way of associational fibers from place cells in the dorsal hippocampus (*22*) and by way of projections from grid cells in the intermediate-to-ventral medial entorhinal cortex, if these exist (*30*). The presence of place cells in the ventral hippocampus suggests that parts of hypothalamus and amygdala involved in motivation and emotion receive significant information about the general location or spatial context of the animal.

The second main finding is that the scale of the hippocampal representation is topographical-

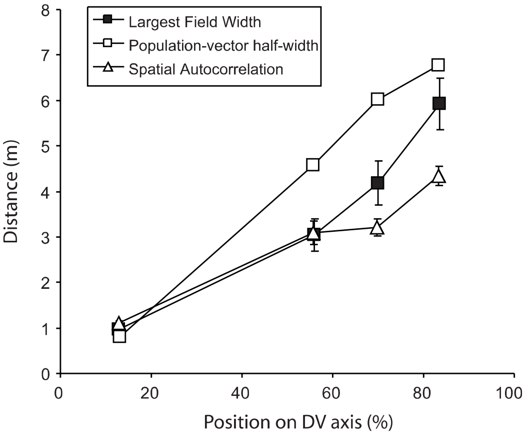


Fig. 4. Spatial scale as a function of position along the dorsoventral (DV) axis of the hippocampus. Scale is expressed as the half-width of the correlated band in the population vector analyses (Fig. 3), as the average width of the largest place field of individual cells (mean ± SEM), or as the estimated field width based on the slope of the best-fit linear regression line for the position versus phase distribution. The population estimate is more likely to reflect the true spatial scale because the algorithm for detecting individual place fields occasionally divided fields as a consequence of local rate variations.

ly graded, much like representations in some sensory maps of the brain (*31*, *32*). The spatial map expands approximately linearly from a scale of <1 m near the dorsal pole to ~10 m near the ventral pole (*33*). Within this range, the hippocampus can represent environments at multiple spatial scales and levels of detail, up to a given limit. The upper limit of ~10 m may be sufficiently large to enable the hippocampus to form detailed representations of all environments within the rat's home range, which has an average radius of about 30 to 50 m (*34*, *35*). A challenge for future research will be to determine whether place cells scale up further in mammals with larger territorial radii.

References and Notes

1. P. Andersen, T. V. P. Bliss, K. K. Skrede, *Exp. Brain Res.* **13**, 222 (1971).
2. M. P. Witter, H. J. Groenewegen, F. H. Lopes da Silva, A. H. M. Lohman, *Prog. Neurobiol.* **33**, 161 (1989).
3. C. L. Dolorfo, D. G. Amaral, *J. Comp. Neurol.* **398**, 49 (1998).
4. K. M. Kerr, K. L. Agster, S. C. Furtak, R. D. Burwell, *Hippocampus* **17**, 697 (2007).
5. S. C. Furtak, S. M. Wei, K. L. Agster, R. D. Burwell, *Hippocampus* **17**, 709 (2007).
6. E. Moser, M.-B. Moser, P. Andersen, *J. Neurosci.* **13**, 3916 (1993).
7. M.-B. Moser, E. I. Moser, *J. Neurosci.* **18**, 7535 (1998).
8. M.-B. Moser, E. I. Moser, *Hippocampus* **8**, 608 (1998).
9. J. O'Keefe, J. Dostrovsky, *Brain Res.* **34**, 171 (1971).
10. E. I. Moser, E. Kropff, M.-B. Moser. *Annu. Rev. Neurosci*. **31**, in press.
11. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
12. M. A. Wilson, B. L. McNaughton, *Science* **261**, 1055 (1993).
13. M. W. Jung, S. I. Wiener, B. L. McNaughton, *J. Neurosci.* **14**, 7347 (1994).
14. A. P. Maurer, S. R. VanRhoads, G. R. Sutherland, P. Lipa, B. L. McNaughton, *Hippocampus* **15**, 841 (2005).
15. M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, M.-B. Moser, *Science* **305**, 1258 (2004).
16. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, *Nature* **436**, 801 (2005).
17. P. Y. Risold, L. W. Swanson, *Science* **272**, 1484 (1996).
18. M. Pikkarainen, S. Rönkkö, V. Savander, R. Insausti, A. Pitkänen, *J. Comp. Neurol.* **403**, 229 (1999).
19. G. D. Petrovich, N. S. Canteras, L. W. Swanson, *Brain Res. Brain Res. Rev.* **38**, 247 (2001).
20. D. M. Bannerman *et al.*, *Behav. Neurosci.* **116**, 884 (2002).
21. K. G. Kjelstrup *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10825 (2002).
22. D. G. Amaral, M. P. Witter, *Neuroscience* **31**, 571 (1989).
23. H. A. Steffenach, R. S. Sloviter, E. I. Moser, M.-B. Moser, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3194 (2002).
24. Materials and methods are available as supporting material on *Science* Online.
25. J. O'Keefe, M. L. Recce, *Hippocampus* **3**, 317 (1993).
26. A. P. Maurer, S. L. Cowen, S. N. Burke, C. A. Barnes, B. L. McNaughton, *Hippocampus* **16**, 785 (2006).
27. M. R. Mehta, A. K. Lee, M. A. Wilson, *Nature* **417**, 741 (2002).
28. K. M. Gothard, W. E. Skaggs, B. L. McNaughton, *J. Neurosci.* **16**, 8027 (1996).
29. Previous studies have demonstrated place fields in intermediate parts of the hippocampus, at approximately 30 to 45% of the dorsal-to-ventral axis (*13*, *14*). Ventral place cells were also reported in another study (*36*) but because this study did not include histological assessment of the recording locations and because the size of the fields was similar to that of dorsal place cells, it is hard to compare the results.
30. T. Solstad, V. H. Brun, K. B. Kjelstrup, M. Fyhn, M. P. Witter, E. I. Moser, M.-B. Moser. *Soc. Neurosci. Abstr.* **33**, 93.2 (2007).
31. T. McLaughlin, D. D. M. O'Leary, *Annu. Rev. Neurosci.* **28**, 327 (2005).
32. J. G. Flanagan, *Curr. Opin. Neurobiol* **16**, 59 (2006).
33. Place fields on left and right runs on the 18-m track were more correlated in ventral CA3 than in dorsal CA3, where representations for different directions are often maximally orthogonalized (*37*) (see supporting online text). This suggests that when a large environment is experienced as connected, as on the track, a single extended representation may be formed in the ventral hippocampus. For discrete environments, ventral CA3 still had strongly orthogonalized place maps (supporting online text).
34. W. B. Jackson, in *Wild Mammals of North America*, J. A. Chapman and G. A. Feldhamer, Eds. (Johns Hopkins Univ. Press, Baltimore, MD, 1982), pp. 1077–1088.
35. D. C. Stroud, *J. Mammal.* **63**, 151 (1982).
36. B. Poucet, C. Thinus-Blanc, R. U. Muller, *Neuroreport* **5**, 2045 (1994).
37. S. Leutgeb, J. K. Leutgeb, A. Treves, M.-B. Moser, E. I. Moser, *Science* **305**, 1295 (2004).

38. We thank R. Skjerpeng for programming; L. Colgin, M. Fyhn, and F. Tuvnes for assistance with experiments or analyses; and A. M. Amundsgård, I. M. F. Hammer, K. Haugen, K. Jenssen, B. H. Solem, and H. Waade for technical assistance. The work was supported by the Centre of Excellence scheme of the Norwegian Research Council and the Kavli Foundation.

Supporting Online Material
www.sciencemag.org/cgi/content/full/321/5885/140/DC1
Materials and Methods
SOM Text
Figs. S1 to S8
References

27 February 2008; accepted 30 April 2008
10.1126/science.1157086

New Products

Spinner Flask

The MatriMix Spinner Flask is a patent-pending, disposable one-liter flask molded from virgin polycarbonate. The automation-friendly flask is fitted with a central magnetic stir paddle that maximizes nutrient flow. A unique teardrop shape breaks up laminar flow and allows for uninterrupted aspiration while spinning. It reduces the risk of cross-contamination by eliminating the multiple wash steps between culture transfers.

MatriCal
For information 509-343-6225
www.matrical.com

Protein Biomarker Verification Platform

MASSterclass is a sensitive protein biomarker verification tool that addresses a key bottleneck in biomarker development by enabling large-scale screening of biomarker candidates without the requirement to raise antibodies. The system improves antibody-free protein assays by quantitating in the low nanogram per milliliter concentration range in blood plasma, thereby enabling efficient and reliable measurement of proteins specific to disease or drug response. The system incorporates a novel series of orthogonal complexity reduction steps combined with targeted tandem mass spectrometry to achieve required sensitivity levels while maintaining reproducibility. The system can screen large panels of biomarker candidates by multiplexing different analytes in the same assay. The new platform complements MASStermind, a biomarker discovery platform capable of generating a high quality hit list of low-abundance protein biomarkers in blood. The combined platforms represent an integrated approach that significantly increases the chances of taking a biomarker from discovery through to clinical implementation within a practical time frame.

Pronota
For information +32-(0)9-241-11-47
www.pronota.com

Cation and Anion Measurement

The Corona CAD universal detector provides a simple and reliable high performance liquid chromatography method for simultaneous measurement of cations and anions. Because cations and anions are key components of everything from pharmaceuticals to food and beverages to fine chemicals, analysis of these atoms and molecules is a critical step in product development and quality control. Measuring both in a single run is more efficient than traditional techniques using dedicated, single-purpose ion chromatography systems. The new method combines the CAD's universal detection with hydrophilic interaction chromatography.

ESA Biosciences
For information +44-1844-239381
www.esainc.com

IgG Assays

The Guava RapidQuant IgG Assays for quantitation of mouse and human antibodies are flexible and easy to use. The assays eliminate the most tedious portions of antibody titering methods. The bead-based assays rely on fluorescence detection, with fluorescent intensity proportional to the concentration of antibody captured on the beads. Standard curves are automatically produced by an easy-to-use software interface, removing the need for time-consuming calculations. The assays have been tested with human and mouse antibodies of all IgG subtypes and in many types of hybridoma media.

Guava Technologies
For information +44-1780-764390
www.guavatechnologies.com

Microwave Synthesizer

The Advancer Kilobatch large-scale automated sequential batch microwave synthesizer is designed for safe and scalable production of up to kilogram quantities of material. It is designed for reliable scale-up without requiring reoptimization, resulting in significant time savings. Unlike traditional microwave batch and flow reactors, the Advancer Kilobatch is capable of processing both heterogeneous and homogeneous reaction mixtures in automated batch-sequencing mode. Safe, reliable, and easy to use, it operates at the high temperatures and pressures required for microwave chemistry.

Biotage
For information +46-18-56-59-00
www.biotage.com

TissueView Software

TissueView Software is designed to improve mass spectrometry (MS) imaging by making data processing more intuitive and delivering easier-to-interpret research results for deeper understanding of how drugs and proteins are distributed within biological tissue. The new software gives researchers a better way to process, visualize, and interpret information that can help pharmaceutical companies select the drug compounds to develop and help academic researchers learn about biological functions. The software rapidly displays the spatial location and intensity of compounds, both proteins and small molecules, in all tissue types. It offers a seamless link between optical images, mass spectral images, MS, and MS/MS spectra.

Applied Biosystems
For information 650-570-6667
www.appliedbiosystems.com/massspectrometry

Electronically submit your new product description or product literature information! Go to **www.sciencemag.org/products/newproducts.dtl** for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.